Author responses to Anonymous Referee #1 are in bold below.

The EPA OTM 33A measurement technique is a mobile inspection method that can provide rapid assessment (~20 minutes) of whether a near-field, near-ground-level source is leaking and at what rate. The method has been widely used to detect and quantify methane emissions from oil and gas production well sites. The method was originally submitted by EPA's Office of Research and Development for inclusion in the Other Test Category (OTM) and is currently in draft form. Several researchers, including EPA's ORD, have previously performed controlled release tests involving single point-source releases to assess the performance of OTM 33A. This study expands on these previous tests by assessing OTM 33A performance under more realistic conditions using a faux oil and gas well site with multiple leak sources from typical well pad equipment. Since the most commonly used OTM 33A emission rate quantification approach (i.e., the point source Gaussian) assumes all emissions from a site converge to a point source, the use of a more realistic test environment with multiple sources provides a means to test the limits of this assumption. The authors' conclusion that, under this more realistic test conditions, OTM 33A has a "small but statistically insignificant low bias" and "does not drastically underestimate total emissions for an ensemble or group of measurements," is supported by the data and the analysis presented here. The paper is well written and the subject matter addressed here is important. However, the authors should consider providing additional details before the paper can be accepted for publication. In particular, the section describing OTM 33A sensitivity to source distances needs to be revised and clarified. Specific comments are provided below.

The authors greatly appreciate the reviewer's carefully considered comments on the manuscript. We have modified the manuscript to address the comments raised by the reviewer and we believe it is much improved because of these adjustments.

Page 2, line 13 to 14. Please expand on or provide a specific reference for the statement that VOC-rich emission sources are difficult to measure with onsite techniques.

A reference to the study by Brantley et al., 2015, which found that high volume samplers could malfunction in VOC-rich emission streams, has been added.

P2, L14-17 now reads: "Drawbacks of onsite measurements include difficulty measuring volatile organic compound (VOC) rich emission sources Brantley et al., 2015, inability to reach all emission sources (such as the tops of free-standing tanks), difficulty measuring intermittent sources, and the time required for each inspection Brantley et al., 2014,Bell et al., 2017, Ravikumar et al., 2018."

Page 2, line 25 to 26. This sentence combines tracer flux method limitations (e.g., measurement distances) with method disadvantages (e.g., tracer flux techniques often require more implementation time than OTM 33A). It might be useful to distinguish between the two. Also, please provide a specific reference for the method limitations/requirements.

The tracer flux method limitations versus disadvantages have been separated. We now include a reference to the Yacovitch et al., 2017 study, which states the TFR method measured 2.5 sites/day. It is now explicitly stated that some of the tracers (acetylene is

common) are flammable. This flammability requires many safety protocols for dealing with acetylene, including emergency shutoff valves and minimum wind speed requirements to prevent pooling near possible ignition sources.

P2, L25-28 now read: "Limitations of TFR include the reliance on downwind roadways of sufficient distance (~0.5 - 2 km) and reliable wind direction (Omara et al., 2018; Roscioli et al., 2015). Drawbacks of using TFR to estimate methane emissions include the amount of time required to estimate emissions from one site (2.5–2.8 sites day-1) (Yacovitch et al., 2017), and the need to transport and release compressed tracer gases (some of which are flammable such as acetylene) near O&G facilities."

Page 4, Section 2.3. The OTM 33A emission rate quantification approach (the point source Gaussian) presented in this section is one of many possible quantification methods for OTM 33A. Other techniques (e.g., backward Lagrangian stochastic models) may have different performances than the PSG approach utilized here.

The reviewer is correct, the following text has been added to clarify, "While several quantification approaches are possible with OTM 33A, the one most commonly employed is an inverse Gaussian approach and this approach is the focus of this manuscript."

P4, L25-26 now reads: "While several quantification approaches are possible with OTM 33A, the one most commonly employed is an inverse Gaussian approach, which is the focus of this manuscript."

Page 4, lines 21 to 22. Please note that EPA considers the method to be more broadly applicable (i.e., not just for emission detection and quantification at point sources).

The following comment has been added to the text.

P4, Line 23 now reads: "OTM 33A is one of the EPA Geospatial Measurement of Air Pollution Remote Emission Quantification (GMAP-REQ) techniques that was designed to observe, characterize, and/or quantify emissions from a variety of sources, though OTM 33A has been used most to measure emissions from O&G operations.

EPA specifically identifies three source assessment modes for OTM 33A: (i) concentration mapping, (ii) source characterization and (iii) emission rate quantification.

The text has been corrected to reflect the three assessment modes.

P4, Line 26-27 now reads: "OTM 33A has three operational parts: concentration mapping, source characterization, and emission rate quantification."

We have also added an additional sentence describing source characterization to P4, L29 -30 which reads: "Source characterization includes observations of temporal variability and emissions composition. If enhancements of methane or other trace gases are detected during downwind transects of a possible source, the laboratory is parked 20–200 m directly downwind within the emission plume to quantify emissions." Page 6, Section 3.1. The description of the OTM 33A test releases should be in Methods section. Similarly, the Methods section should include an overview of statistical tests performed, which are described in later sections under Results.

Section 3.1 has been moved to the methods. While description of the statistical tests would typically go in the Methods section, for this manuscript we feel it is appropriate to leave it in the results because the potential bias of certain statistical analyses is a key result.

Page 6, lines 17 to 18. Please spell out how many "multiple release points" there were.

Information regarding the number of unique release points used for the METEC test releases has been added.

P6, L now reads: "For this study, we used one METEC site representative of a small O&G facility that included a condensate storage tank, separator, and well head, all of which were plumbed to be possible emission sources, 11 of which were used in this study(Fig. 1). This resulted in 15 release configurations that had from 1–3 release points at different heights (0.33–4 meters), up to 6 meters apart from one another."

The caption for Figure 1 has also been adjusted to read: "METEC facility with nine of the 11 release points circled. Release points include (clockwise from top of tank) tank candy cane, tank thief hatch, tank front flange, wellhead Kimray packing, wellhead hand valve packing, separator burner fuel supply, separator Kimray vent, separator PRV, and separator house PRV. Not pictured: wellhead lubricator flange and wellhead pressure gauge. The UW mobile laboratory is in the background."

Page 6, lines 21 to 22. What informed the choice for the emission range tested here? Were the authors limited to this range? This has potential implications for how broadly applicable the results are, especially when larger emission rates (beyond the ~ 2 kg/h rate) are encountered in the field.

The release rate range was constrained by the facilities/test release configurations. We would have liked to measure larger release rates, but this was not possible. This information, as well as the range of bootstrapped mean emission rates for four basins and a discussion of the limitations of this measurement range have been added to the Conclusions.

P12, L25-29 : "For both test release experiments, the maximum release rates (2-2.15 kg h⁻¹) were constrained by available resources and facility throughput and, while they represent a large fraction of emission rates observed in the field, they do not fully encompass the dynamic range of emissions observed in an O&G basin. The bootstrapped mean emission rates from four O&G basins measured by the University of Wyoming range from 0.68–3.7 kg h-1(Robertson et al., 2017), suggesting the range of these test releases may not be representative of the largest emission rates observed in the field (Fig. 13)."

Page 6, lines 23 to 24. Did the authors perform tests at different source-to-observation distance configurations? If so, it would be helpful to provide a range/basic statistics here. Additional comment on this below.

The mean and range of measurement distances for each test release experiment has been added to their descriptions.

P6, L16-17: "Mean measurement distance was 78 m, with a range of 34–174 m." P6, L29: "Mean measurement distance was 114 m with a range of 53–195 m."

Additionally, an excel workbook giving the distance to source, release rate, release height, average wind speed, and OTM estimated emission rate has been added to the SI for both test release experiments.

Page 9, Section 3.3.1. This is an important section. Unfortunately, important details are missing. What was the average source distance for all test releases and how does this compare to the average in the Bell et al. study and in the EPA test? The data is shown in Figure 13, but it would be helpful to describe it here. Were there any measurements that were repeated at different source-to-observation distances to test OTM 33A sensitivity to source distances?

Section 3.3.1 (now Section 3.2.1) now includes a summary of average and range of source distance for the test releases and the Arkansas data. We could find no reference for the distance observed for the complete set of EPA test releases (N = 107), but the preliminary report on OTM 33A by Thoma et al., 2012 included 24 test releases with measurement distances ranging from 18-103m.

P10, L9-16 now read: "OTM 33A sensitivity to distance was also tested in the field during the METEC test releases. For configurations that had both a "closer" (generally<70 m) and "farther" (generally>100m) measurement distance for replicate measurements, the closer measurement had a flux estimate closer to the known release 78% of the time (SI Sect. 1.4). The average distance of the closer replicate measurements (78 m) is comparable to the average measurement distances for the CF-TR of 78 m, smaller than the mean METEC-TR distance of 114 m, and larger than the measurement distances during the Arkansas campaign of 46 m (20–113 m) (Robertson et al., 2017; Bell et al., 2017). For both the CF-TR and the METEC-TR, there is no obvious increase in% error as measurement distance increases (Fig. 10(a)), suggesting the underestimation reported by Bell et al. cannot be blamed solely on closer measurement distances."

The METEC test release were designed to measure the same release configuration (release points and release rate) two to three times. Unfortunately, only 10 of the 15 attempted configurations had duplicate measurements that passed the DQI. Initially, we felt the sample size was too small to include statistical tests for this sub-set of data, but we have added this information to Section 3.1.1, 3.2.1, and SI Sect. 1.4.

Lastly, the mean statistics for measured basins and test releases have been added to the caption of Figure 13.

The caption now reads: "Figure 13. Summary of accepted OTM 33A measurements from field deployments and test releases (right of vertical line). Basins from Robertson et al. 2017. Upper Green River Basin, Wyoming (UGRB), Uintah Basin, UT (UB), Denver-Julesburg Basin, CO (DJ), Fayetteville, Arkansas (AR). Mean statistics (from left to right) are as follows. Distance [m]: 98, 114, 83, 51, 114, 78. Flux [kg h-1]: 2.41, 6.99, 1.51, 1.27,0.51, 0.96. Mean wind speed [m s-1]: 5.3, 4.2, 3.1, 2.9, 4.1, 4.9. Stability class: 5.0, 4.9, 5.3, 3.5, 5.0, 5.4."

The second paragraph also needs more clarity. There is ambiguity in how the % changes in source distances were calculated. The % change could be based on (i) measurements of an emission source(s) at different observation distances, which means several 20-min samples of one known release were measured at different observation distances spanning a range of 20 m to 200 m, or (ii) fixed observation location, but the source-to-observation distance is varied (post-measurement) based on whether one assumes an average distance for all onsite sources or distance from a single known point source onsite. In the latter scenario, the difference in source distance would be no more than 6 m, the maximum separation distance for the multiple sources onsite (page 6, lines 17âA~T18), which is small in the context of the 20â ~ A~T200 ~ m range. Please provide more details in this section to help the reader understand how the variability in source distances was assessed. Also, a plot showing the OTM % error as a function of source distance (similar to Figure 10 for wind speed) may help illustrate the point.

More clarification has been added to Section 3.2.1 to distinguish between the two potential source distance errors identified by the reviewer.

Beginning on P9, L25: "OTM 33A sensitivity to source distance was tested two ways for the METEC test releases. The following test was performed during the data analysis stage, and compared the flux estimated using the average distance of all the components that could be sighted with the range finder from the van (e.g. wellhead, separator, tank) to the flux estimated using the distance to the known release point or point distance (identified using the FLIR camera). Although the well pad measured at the METEC facility was quite small (~6 m by 6 m), the average source distance was larger than the specific source distance $\langle \Delta D \rangle$ of the time. The change in the OTM 33A flux ($\Delta Flux$) as a result of changing the measurement distance ($\Delta D \rangle$ was found using Equations 3 and 4.

△Distance=(Average Distance-Point Distance)/(Average Distance) × 100	(3)
% ΔFlux=(Average OTM-Point OTM)/Average OTM × 100	(4)

A correlation plot of % Δ Distance and % Δ Flux suggests that for a 5% change in source distance, the OTM 33A flux estimate would increase by almost 10% (Fig. 9(a)). In terms of mass error, the OTM flux estimated by the average or specific source distance has very little impact in the over- or under-estimation of the METEC known release (Fig. 9(b)). Allowing this fit to have an intercept changes the linear fit toy= 0.978x-0.03, a negligible difference. Source distance related error is small in the context of the ±70% measurement error, but this analysis underscores how determination of the exact emission point can further reduce errors in the field."

Additionally, a panel has been added to Figure 10 to show how estimated fluxes varied versus distance of the mobile lab from the emission source.

Page 9, Section 3.3.3. It is not clear here whether the sampling probe height was fixed or adjusted for different measurements. In Section 2.1 the sample inlet on the mobile laboratory is described to be located "4 meters" above the ground. And source heights are described to vary from 0âA^{T3} meters above the ground (page 6, lines 17â^AT18). In ^{testing} OTM 33A sensitivity to source heights, were there specific configurations where one or more source heights was/were greater than the sampling probe height?

A better description of the mast has been added to Section 2.1 to emphasize that it remains at a fixed height of 4 m.

P4, L5-7 now read: "The University of Wyoming mobile laboratory is a customized Freightliner Sprinter van. The front of the van is equipped with a horizontal mast that projects instrumentation and the inlet at a fixed height of 4 meters above the ground slightly beyond the vehicle's front bumper."

The METEC facility had release points ranging from 0.33-4.33 meters above ground level.

P6, L23-24 now read: "This resulted in 15 release configurations that had from 1–3 release points at different heights (0.33–4 meters), up to 6 meters apart from one another."

Page 10, lines 13 to 14. This is partly correct. The Alvarez et al. study also used other datasets obtained using other measurement techniques, not just the OTM 33A measurements in the four O&G basins described here.

This wording has been corrected. It now states that data from these basins, "represented a significant fraction of data, along with other field campaigns."

Page 16, Figure 2. Please increase the font size for both figures (on all axis labels, legend and tick labels)

Font size has been increased.

Page 25, Figure 11. It is not clear what heights (or range of heights) correspond to the height ranks shown here. What is the highest emission point? What are low, medium and high release points?

Figure 11 has been remade to show the average height of the emission sources in meters instead of by rank.

Page 27, Figure 13. It may be helpful to add a vertical line separating the dataset for actual OTM 33A measurements and OTM 33A release trials.

The suggested change has been made.

General comment on all figures: some figures have figure titles and others do not. Please review AMT guidelines and revise accordingly.

All figures now have figure titles.

Author responses to Anonymous Referee #2 are in bold below:

This paper deals with a commonly used ground-based method (OTM33A) for estimating emissions rates. Recent papers have highlighted the relevance of site-level (facility-wide) emission estimates. The authors perform tests to assess accuracy of this approach in the context of methane emissions from single sites as well as ensembles (i.e., characterize emissions distributions from a population of sites). The results are relevant due to the increasing use of the approach. I recommend publication after some minor edits/clarification. Two main points to be addressed/expanded by the authors: (1) Effect of multiple sources-distance selection (2) Determination of non-detects and potential effect of overestimation in determining fraction of sites that fall below detection limit. Additional comments:

We greatly appreciate the reviewer's careful consideration of the manuscript. We have addressed the noted issues as detailed below.

INTRODUCTION: Page 1, Line 23: "Site-level measurements are therefore necessary for improving emission estimates of the O&G production sector." This is true, might be also useful to mention importance of site-level measurements in conjunction with component-level measurements to understand source of emissions.

The text on line 23 has been changed to include component-level measurements and the study by Brandt et al., 2014, is included to help emphasize that point.

P1, L23-24 now read: "Site- and component-level measurements are therefore necessary for improving emission estimates of the O&G production sector (Brandt et al.,2014).

Page 2, Line 3: 'However, more permanent approaches are still under development and must be approved as equivalent monitoring technologies before they can replace existing EPA approved Leak Detection and Repair (LDAR) methods like optical gas imaging (OGI).' Suggest expanding discussion of difference between leak detection and leak (emissions) quantification, which is important in the context of LDAR and equivalency. One could argue that main goal of LDAR is not improving inventories, but repairing leaks. I think this idea needs to be further developed to link it to importance of site-level measurements.

Additional text has been added to page in an attempt to emphasize that LDAR does not typically generate data that can be used to improve emission inventories. While we agree that there is ample material to be discussed in terms of LDAR methods and equivalency, we believe this is beyond the scope of the current manuscript.

P2, L6-10 now read: "Annual or semi-annual LDAR programs already in place rarely quantify total emissions from a site, and the efficacy of these programs depends on many factors including employee experience, leak size, and meteorological variables like wind speed and temperature (Ravikumar et al., 2016, 2018). This makes LDAR programs an important tool for finding leaks and reducing emissions, but they often do not explicitly quantify or provide data of the actual emission rate from production sites, and this limits usefulness for improving emission inventories."

METHODS It might be useful to briefly discuss the detection limit of the method (threshold for considering a site as non-detect). This is discussed in previous papers, but might be useful to summarize here. Consequently, discuss the potential overestimation at lower emission rates with the threshold for non-detects. This is something that matters for the ensemble.

The method limit of detection has been added to the methods section. P5, L30-31 now reads: "The estimated lower detection limit of the method is 0.01 g s⁻¹ 0.036 kg h⁻¹) (Brantley et al., 2014)."

In the current study there were no "non-detects" meaning there is no bias in the Christman or METEC ensembles. In the field, careful consideration of non-detects is essential, but we feel this is best addressed in the papers covering those field deployments as methodology varies slightly from one study to another. Overall, the slight underestimation of total mass flux found in this study and the large underestimation reported by Bell et al., 2017 support OTM 33A being, if anything, slightly low for an ensemble of measurements. In general, it is not too critical what number is inserted for the low emission wells as the mean of the ensemble is dominated by higher emission sites and the uncertainty in the number of high emission sites.

Page 5, line 23: Might be good to mention that this could also affect flares (in addition to liquids unloadings).

P5, L27-29 now read: "OTM 33Astruggles to quantify plumes with a particularly high vertical velocity or buoyancy (such as manual unloadings, lit or unlit flares, or very hot emissions)."

Page 9, line 11-14. What happens with multiple sources on site? This paragraph hints at the importance of using OGI to locate source. Might be useful to expand on distance selection under various sources (i.e., based on highest emission point?)

The analysis presented here suggests that, at least for emission points that are within 6 m of each other, no selection of a specific source is necessary given that the observed error of ~10% is much smaller than other errors associated with the method. This section has been expanded to more clearly explain the relatively small impact of not knowing the exact source location on smaller sites (these are the sites typically measured via OTM 33A).

Page 10, line 9-11. 'These results also indicate OTM 33A does not drastically underestimate the total emissions for an ensemble or group of measurements, and that scaling up mean emissions measured with OTM 33A to an entire basin is a valid approach." This is an important conclusion from the paper since the ensemble is a common application of this method. Might be good idea to further highlight in the abstract.

We agree this is an important finding. We believe the statement in the abstract on Page 1 Lines 12-13 that, "an ensemble of OTM 33A measurements may have a small but statistically insignificant low bias." makes this point without overstating what can be determined from the current study.

Figure1: It might be useful to expand caption to include label of release points (i.e., what is the source of emissions).

The caption of Figure 1 has been expanded to include descriptions of all of the pictured release points, as well as the total number of release points (11).

Figure 2: Significant figures for R parameter. Significant figures for all R parameter values have been appropriately reduced.

Constraining the Accuracy of Flux Estimates Using OTM 33A

Rachel Edie¹, Anna M. Robertson¹, Robert A. Field¹, Jeffrey Soltis¹, Dustin A. Snare², Daniel Zimmerle³, Clay S. Bell³, Timothy L. Vaughn³, and Shane M. Murphy¹

¹University of Wyoming 1000 E. University Ave. Laramie, WY 82070

²All4 Inc., Kimberton, Pennsylvania 19442, United States

³Colorado State University Energy Institute, 430 N College Ave. Fort Collins, CO 80524

Correspondence: Shane M. Murphy (shane.murphy@uwyo.edu)

Abstract.

Other Test Method 33A (OTM 33A) is a near-source flux measurement method developed by the Environmental Protection Agency (EPA) primarily used to locate and estimate emission fluxes of methane from oil and gas (O&G) production facilities without requiring site access. A recent <u>nation-wide_national</u> estimate of methane emissions from O&G production included

- 5 a large number of flux measurements of upstream O&G facilities made using OTM 33A and concluded the EPA National Emission Inventory underestimates this sector by a factor of ~2.1 (Alvarez et al., 2018). The study presented here investigates the accuracy of OTM 33A through a series of test releases performed at the Methane Emissions Technology Evaluation Center (METEC), a facility designed to allow quantified amounts of natural gas to be released from decommissioned O&G equipment to simulate emissions from real facilities (Fig. 1). This study includes test releases from single and multiple points, from
- 10 equipment locations at different heights, and spanned methane release rates ranging from 0.16 to 2.15 kg h⁻¹. Approximately 95% of individual measurements (N=45) fell within \pm 70% of the known release rate. A simple linear regression of OTM 33A versus known release rates at the METEC site gives an average slope of 0.96 with 95% CI (0.66,1.28), suggesting that an ensemble of OTM 33A measurements may have a small but statistically insignificant low bias.

1 Introduction

- 15 Methane is a potent greenhouse gas, and emissions from the oil and gas (O&G) sector are thought to account for roughly 30% of total methane emissions in the United States (U.S. EPA, 2019). "Upstream" O&G activities (extraction, production, etc.) are thought to contribute the bulk of emissions within the O&G sector (Alvarez et al., 2012; Zavala-Araiza et al., 2015; Alvarez et al., 2018). However, attempts to quantify O&G methane emissions are hindered by inaccurate emission inventories, a lack of measurements, and variability between basins (Allen, 2016; Schwietzke et al., 2017; Robertson et al., 2017; Alvarez et al.,
- 20 2018; Omara et al., 2018). For example, basin-wide aircraft measurements of methane emissions from different O&G basins find emissions are generally higher than official inventories published by the U.S. Environmental Protection Agency (EPA) (e.g., Karion et al., 2013; Pétron et al., 2014; Karion et al., 2015; Schwietzke et al., 2017; Peischl et al., 2018)

(e.g., Karion et al., 2013; Pétron et al., 2014; Brandt et al., 2014; Karion et al., 2015; Schwietzke et al., 2017; Peischl et al., 2018), but the scale of aircraft measurements give little insight into the exact source of emissions on the ground. Site-level Site-

and component-level measurements are therefore necessary for improving emission estimates of the O&G production sector (Brandt et al., 2014). Existing near-source studies of O&G basins suggest the majority of large, uncontrolled emissions are the result of faulty equipment that may not be noticed for some time (Zavala-Araiza et al., 2017; Omara et al., 2018), emphasizing the need for permanent or semi-permanent monitoring technologies instead of infrequent manual inspections (Coburn

- 5 et al., 2018; van Kessel et al., 2018). However, more permanent approaches are still under development and must be approved as equivalent monitoring technologies before they can replace existing EPA-approved Leak Detection and Repair (LDAR) methods like optical gas imaging (OGI). Additionally, annual <u>Annual</u> or semi-annual LDAR programs already in place rarely quantify total emissions from a site, and the efficacy of these programs depends on many factors including employee experience, leak size, and meteorological variables like wind speed and temperature (Ravikumar et al., 2016, 2018). This makes
- 10 LDAR programs an important tool for finding leaks and reducing emissions, but they often do not explicitly quantify or provide data of the actual leak emission rate from production sites, and this limits usefulness for improving emission inventories. In the absence of OGI-equivalent continuous monitoring approaches, both basin and site-level emission estimates have been gathered using a number of different techniques, all with strengths and weaknesses.
- One approach is to measure emissions onsite of an O&G production facility. Onsite measurement teams typically detect emissions from malfunctioning components via OGI, which can be quantified using high-volume samplers. Drawbacks of onsite measurements include difficulty measuring volatile organic compound (VOC) rich emission sources (Brantley et al., 2015) , inability to reach all emission sources (such as the tops of free-standing tanks), difficulty measuring intermittent sources, and the time required for each inspection (Brantley et al., 2014; Bell et al., 2017; Ravikumar et al., 2018). Site access requirements also introduce the possibility of changes in operation when measurement teams are onsite (Alvarez et al., 2018).
- The tracer flux ratio (TFR) technique estimates methane emissions by multiplying the observed concentration ratio of methane to a tracer by the known emission rate of the tracer. TFR has been used in both ground-based (Roscioli et al., 2015; Yacovitch et al., 2015) and airborne applications (Daube et al., 2018), though only ground-based approaches have been used for O&G facilities. TFR can quantify all emissions at an O&G facility, and can often differentiate between emissions from relatively close facilities without the need for site access, though access can improve flux estimates (Roscioli et al., 2015)
- 25 . TFR does not require an atmospheric transport model and is therefore insensitive to uncertainties in atmospheric stability and turbulence. Drawbacks of this technique Limitations of TFR include the reliance on downwind roadways of sufficient distance (~0.5 2 km) , and reliable wind direction , (Omara et al., 2018; Roscioli et al., 2015). Drawbacks of using TFR to estimate methane emissions include the amount of time required to estimate emissions from one site (>1 hr) 2.5–2.8 sites day⁻¹) (Yacovitch et al., 2017), and the need to release transport and release compressed tracer gases (typically acetyleneor nitrous
- 30 oxidesome of which are flammable such as acetylene) near O&G facilities.

As mentioned previously, airborne mass flux measurements have been used to estimate methane emissions from multiple O&G basins (eg., Karion et al., 2013, 2015; Peischl et al., 2015, 2016, 2018; Pétron et al., 2014; Schwietzke et al., 2017). Meteorological requirements (like a fully developed planetary boundary layer and consistent wind direction) make these measurements difficult, especially for expansive O&G basins such as the Permian basin in Texas and New Mexico (Peischl et al., 2017).

35 2018). Emission estimates of individual production sites via aircraft measurements are also possible, but measurement sites

typically need to have relatively large emissions and are limited by aircraft range, turning radius, and favorable meteorological conditions (Caulton et al., 2014; Lavoie et al., 2015, 2017; Conley et al., 2017). Additionally, airborne sampling must occur during the day to meet meteorological requirements, and diurnal variability of emissions associated with onsite maintenance could impact aircraft-based emission estimates in some basins (Schwietzke et al., 2017; Vaughn et al., 2018; Zaimes et al., 2019).

5

A final type of measurement technique used to estimate emissions from O&G production facilities—and the focus of this study—are downwind measurements that estimate emissions by using methane mixing ratio and wind measurements to derive the source flux. Downwind emission flux estimates are made using parameters measured in the field combined with additional parameters found with Gaussian or atmospheric dispersion models (Brantley et al., 2014; Rella et al., 2015; Caulton et al.,

- 10 2018; Robertson et al., 2017; Lan et al., 2015; Foster-Wittig et al., 2015). Downwind measurements do not require site access, but may not be able to identify or capture all sources onsite, especially buoyant ones. Similar to TFR, these techniques require downwind roadways (50 200 m away) and consistent wind direction. Operator-approved site access can improve OTM 33A measurement success in regions with limited downwind roadway infrastructure or complex topography. Though sampling time can be considerably faster than TFR or onsite techniques, it is hard to measure enough sites to get a representative sample
- 15 (and therefore a flux) of an entire O&G basin (Harriss et al., 2015). As a whole, all of the emission measurement techniques mentioned here are only representative of a timescale between seconds and hours, and therefore have difficulty capturing emissions sources with large temporal variability (U.S. EPA, 2014; Brantley et al., 2014; Robertson et al., 2017; Bell et al., 2017; Caulton et al., 2018; Vaughn et al., 2018).
- This study focuses on a ground-based mobile emissions measurement approach, Other Test Method 33A (OTM 33A). OTM 33A is among the most common downwind methods, along with TFR, used to measure methane and VOC fluxes from O&G sources (Brantley et al., 2014, 2015; Robertson et al., 2017). A recent study by Bell et al. (2017) compared onsite, OTM 33A, and TFR measurement techniques in the Fayetteville Shale. The results of the Bell et al. study suggest OTM 33A only captured ~40-60% of emissions measured or estimated by onsite teams in the Fayetteville when the dominant emission source was an onsite direct measurement rather than a simulated emission source. OTM 33A had a larger low bias when manual or automated
- 25 unloadings were measured. Manual or automated unloadings occur when the well pressure is not great enough to move liquids from the geologic formation, preventing gas flow to the pressurized sales line. To maximize the pressure differential, the well is vented directly to the atmosphere in order to remove accumulated liquids. This process can be performed manually or automatically, and may use a plunger to assist with liquid removal. This creates an emissions plume with high vertical velocity. It is likely the majority of this plume would pass over the mobile laboratory unless perfect conditions and road access
- 30 generate a downwind measurement site 200 m or less from the source. The results of the Bell et al. study add uncertainty to recent national methane emission estimates, which relied heavily on OTM 33A measurements in five O&G basins (Alvarez et al., 2018). However, the Alvarez et al. study also found that basin-wide emission estimates based on OTM 33A facility measurements agreed with airborne basin-wide flux estimates to within measurement uncertainty. Additionally, no significant low-bias (>10%) was detected in numerous (>100) OTM 33A test releases, conducted by multiple groups (Brantley et al., 2014;
- 35 Robertson et al., 2017). These test releases were all single point-source releases conducted in open terrain without obstacles,

which may not be a reliable comparison to the types of emission sources experienced in O&G fields. The discrepancy between results of Bell et al. study and previous test releases, along with the potential significant impact on national emission estimates, motivated the suite of more realistic test releases described here.

2 Materials and Methods

5 2.1 Mobile Laboratory

The University of Wyoming mobile laboratory is a customized Freightliner Sprinter van. The front of the van is equipped with a <u>horizontal</u> mast that projects instrumentation and the inlet at a fixed height of 4 meters above the ground slightly beyond the vehicle's front bumper. Meteorological instruments on the mast include a 3-D sonic anemometer and an all-in-one compact weather station. The mast also includes a camera, an AirMar differential GPS, and a Teflon inlet (1/4") OD for

- 10 gas-phase species. Ambient air is pulled through the Teflon inlet at a rate of 6.5 L min⁻¹. For the test releases described here, the laboratory was instrumented with a G2204 Picarro Cavity Ringdown Spectrometer (CRDS) which has been modified to measure water vapor and dry methane concentrations at a frequency of 2 Hz. The Picarro has an additional meter of 1/8" OD Teflon tubing that branches from the main inlet line, resulting in a total sample transit time through the inlet to the instrument of one second. This lag is accounted for during data processing. Additionally, the van contains a battery bank which allows the
- 15 instrumentation and data acquisitions system to be used while the vehicle engine is turned off.

2.2 Instrument Calibration

The Picarro response was tested using two NIST certified methane-zero air mixtures (2.538 ± 0.05 ppm, 101 ± 5 ppm), and ultra-high-purity zero air (UHPA) at intervals throughout the campaign to confirm stability and accuracy. The instrument was always within ± 0.01 ppm of the lower NIST standard, ± 1 ppm of the higher standard, and ± 0.003 ppm of zero when tested

with UHPA. The 5-second instrument precision is \pm 0.002 ppm. Due to the observed instrument stability and accuracy, no calibration adjustments were made to methane concentrations during data processing.

2.3 OTM 33A Measurement Method

OTM 33A is one of the EPA Geospatial Measurement of Air Pollution Remote Emission Quantification (GMAP-REQ) techniques that was designed to observe, characterize, and/or quantify emissions from a variety of sources, though OTM 33A has

- 25 been used most often to measure emissions from O&G operations (U.S. EPA, 2014; Thoma, 2012; Brantley et al., 2014, 2015; Robertson et al., 2017). While several quantification approaches are possible with OTM 33Ahas two operational parts; first to detect and second to quantify emissions, the one most commonly employed is an inverse Gaussian approach, which is the focus of this manuscript. OTM 33A has three operational parts: concentration mapping, source characterization, and emission rate quantification. Detection of emissions occurs by driving downwind of possible emission sources in an attempt to transect
- 30 an emissions plume, measure the ambient background trace gas mixing ratio, and, if possible, to rule out any emissions from

upwind sources. <u>Source characterization includes observations of temporal variability and emissions composition</u>. If enhancements of methane or other trace gases are detected during downwind transects of a possible source, the laboratory is parked 20–200 m directly downwind within the emission plume to quantify emissions. Care is taken to <u>orientate position</u> the mast directly into the dominant wind direction to minimize impact from turbulent eddies around the vehicle. Once the laboratory is

- 5 safely positioned, the vehicle is turned off and an OTM 33A flux measurement begins. During the ~20 minute measurement, 2 Hz measurements of wind direction (in x, y, and z), wind speed, temperature, and the methane mixing ratio are collected and time-stamped with a universal data system time. Meanwhile, distance to the possible emission sources relative to the mast of the laboratory are measured using a TruePulse laser range finder (Model 200). If possible, the most likely emission source is identified using an infrared camera (FLIR GF300). Site photos and observations are also collected.
- 10 The OTM 33A analysis program, written in MATLAB (2015), estimates an emission mass flux, Q [g s⁻¹], by using the Gaussian dispersion equation (Eq. 1). The terms of this equation are found as follows. First, the lowest 5% of measured mixing ratios during the \sim 20 minute measurement are averaged and considered ambient background, which was around 1.9 ppm (±0.15 ppm) of methane for this study. The background value is subtracted from the data to yield methane enhancement. The analysis program bins observed methane enhancements by wind direction into 10° bins (Fig. 2(a)), and then calculates the
- 15 average methane enhancement observed in that wind bin. A plot of methane enhancement vs. wind direction is then generated and fit to a Gaussian distribution (Fig. 2(b)). The Gaussian fit's apex is C_{peak} [g m⁻³]. To determine the expected spreading of the emission plume, the program calculates atmospheric stability indicator values (ASI). The ASI are based on the standard deviation of the 2-dimensional wind direction (horizontal spreading), and the standard deviation in vertical wind speed (vertical spreading), also known as the turbulent intensity. The horizontal and vertical ASI are averaged together into a Point Gaussian
- 20 Indicator (PGI) value, which parameterizes the vertical and horizontal plume spread experienced during the OTM measurement. There are seven PGI values which correspond to Pasquill stability classes A-D (Brantley et al., 2014; U.S. EPA, 2019). The PGI and measured source distance are used as inputs to a lookup table that gives the plume dispersion in two dimensions, σ_y [m] and σ_z [m]. The average wind speed \overline{U} [m s⁻¹] is also calculated for the same time periods methane enhancements are observed.

25
$$Q = 2 \times \pi \times \sigma_y \times \sigma_z \times \overline{U} \times C_{\text{peak}}$$
 (1)

Equation 1 does not include any terms for ground reflection of the plume, plume buoyancy/velocity, or differences in height of the emission source and measurement inlet. OTM 33A assumes a single emission point. For this reason, OTM 33A is best suited for measuring O&G facilities with equipment concentrated in one area that have downwind roadways. OTM 33A struggles to quantify plumes with a particularly high vertical velocity or buoyancy (like manual unloadingsor such as manual

30 <u>unloadings</u>, lit or unlit flares, or very hot emissions). In this scenario, the calculated C_{peak} will not represent the center of the emission plume, leading to underestimations of these sources (Bell et al., 2017). The estimated lower detection limit of the method is 0.01 g s⁻¹ (0.036 kg h⁻¹) (Brantley et al., 2014).

A series of built-in data quality indicators (DQI) will flag an OTM 33A flux estimate for a variety of reasons, including poor Gaussian fit, inadequate sampling time within the emission plume, too variable wind speed or direction, or a maximum

methane enhancement that is too small. Flags are then added up, and measurements are broken into categories that represent the probability an OTM measurement is a good flux estimate. For the current study, the same approach as Robertson et al. (2017) and Bell et al. (2017) was used where most of the Category 1 and a few Category 2 measurements that were only flagged for low methane concentrations (max enhancement less than 100 ppb above background) were considered. Occasionally, measurements

5 with very few DQI flags (Category 1 measurements) will be thrown out after review of the Gaussian fit or if IR camera images suggest we are missing most of the emission plume. Full descriptions of the DQI can be found in SI Sect. 1.2, Robertson et al. (2017), Brantley et al. (2014), and in the EPA's documentation (U.S. EPA, 2014).

3 Results

2.1 Test Releases

- 10 The University of Wyoming performed two sets of test releases to assess the ability of OTM 33A to quantify methane emissions. The first set of tests, the Christman Field Test Releases (CF-TR) were conducted in conjunction with Colorado State University in July and August of 2014 at the abandoned Christman Airfield in Fort Collins, CO. These releases consisted of two configurations, a simple point source (an opened gas cylinder) and manifold (an elevated ~6-foot length of PVC pipe with many perforations). Neither source of methane gas was obstructed, and they were, in essence, single point sources, one slightly
- 15 broader than the other. Release rates were set using calibrated mass flow controllers and are correct to within 5%. These tests spanned a variety of release rates (0.2 to 2 kg hr⁻¹) and were staged in an open field with no obstructions (clear line of site) between the single methane source and mobile lab. Winds ranged from 2–8 m s⁻¹ from the S/SE. The calculated PGI ranged from 2–6, which roughly correspond to Pasquill-Gifford stability classes A–D. Mean measurement distance was 78 m, with a range of 34–174 m. Details of these results are reported in Snare (2015) and Robertson et al. (2017).
- 20 The more-recent set of tests were performed at the Methane Emissions Technology Evaluation Center (METEC) in Fort Collins, CO in June of 2017. METEC contains multiple faux O&G facilities ranging in size and complexity with decommissioned O&G equipment that has been plumbed to release a known amount of natural gas (>94% methane) from a multitude of points. For this study, we used one METEC site representative of a small O&G facility that included a condensate storage tank, separator, and well head, all of which were plumbed to be possible emission sources, 11 of which were used in this study (Fig.
- 1). This resulted in 15 release configurations that had multiple from 1–3 release points at different heights (0-3-0.33-4 meters), and up to 6 meters apart from one another. The relative complexity of the site also introduced obstructions (the methane release would have to flow around a large tank or other piece of equipment to reach the mobile lab) which could potentially impact release quantification. Releases spanned 0.17 to 2.15 kg hr⁻¹ and were controlled by combining flows from a number of critical orifices, resulting in a four σ release error less than 5%. Meteorological conditions ranged from sunny to partly cloudy, with
- 30 average winds from 2–9 m s⁻¹ from the E/SE. The calculated PGI ranged from 3–6, which roughly correspond to Pasquill-Gifford stability class B–D. Mean measurement distance was 114 m with a range of 53–195 m. One to two duplicate OTM 33A measurements were attempted at different distances for each of the 15 unique METEC test releases configurations.

3 Results

23 OTM 33A test releases were measured during the CF-TR; 21 passed the data quality indicators (DQI) (U.S. EPA, 2014) and were included in this analysis. 34 test releases were measured during the METEC-TR, of which 24 passed the DQI and were included in this study. A similar success rate, \sim 70%, has been observed in the majority of the basins measured by the

5 University of Wyoming (Robertson et al., 2017). Of the 24 successful measurements during the METEC test releases, there were 10 replicate measurements (same release configuration but different OTM 33A measurement distances). The following analysis explores different statistical approaches to constrain the error associated with individual OTM 33A measurements and to assess the accuracy and precision of an ensemble of OTM 33A measurements. The latter analysis is especially important given this is how OTM 33A measurements are often scaled up to estimate basin-wide emissions from O&G.

10 3.1 Evaluating the Accuracy of OTM 33A

3.1.1 Percent Error Analysis

$$\% Error = \frac{OTM\,33A\,flux - known\,release}{known\,release} \times 100\tag{2}$$

Percent error, Equation 2, Percent error (Eq. 2) was calculated for each individual measurement made during the test releases. A histogram of percent error for both the CF-TR and METEC-TR indicate a large range in over- and under-estimations

- 15 are possible using OTM 33A (Fig. 3). Percent error ranges from -75% to 50% and -60% to 170% for CF-TR and METEC-TR respectively. Figures 4 and 5 show that the larger % errors correlate with smaller release rates, with OTM 33A generally overestimating smaller releases. 68% of the CF-TR data fall within $\pm 28\%$ of the known release, which is the 1 σ error used by Robertson et al. and similar to the error reported by the EPA of 72% of measurements within $\pm 30\%$ of the known release (Brantley et al., 2014). 68% of the METEC-TR are within $\pm 38\%$ of the known release, perhaps suggesting that a slightly higher
- 20 1σ error is appropriate, especially if measuring emissions fluxes less than 0.5 kg hr⁻¹. For the combined set of test releases, greater than 85% of the data are within ±50% of the known value, and 95% of the data are within ±73%. If a Gaussian curve is fit to all of the test release data (N=45), the 95% confidence interval is found to be +54% to -84%, suggesting a low bias of -15% and a 2σ error of ±69% (Fig. S5). The rounded 2σ confidence interval for test releases of ±70% would become 0.58*q* and 3.33*q* when *q* is an OTM 33A estimate made of an unknown emission source in an O&G basin. The number of replicate
- 25 measurements of METEC release configurations were too small to perform a similar statistical analysis (N=10), but multiple measurements did not decreased the mean OTM 33A measurement error (14.7% for replicate measurements, 13.1% for all measurements). Replicate measurements have been shown to improve flux estimates but at the expense of measuring a number of unique sites (Brantley et al., 2014).

3.1.2 Ordinary Least Squares Regression

30 Another approach to assess the performance of OTM 33A is using an ordinary least squares (OLS) regression applied to a correlation plot of the OTM 33A flux estimate versus the known release rate. Assuming the OTM-measured flux and known

release rate converge at (0,0) yields OLS slopes of 0.91 for CF-TR and 0.92 for METEC-TR (Fig. 6). This suggests OTM 33A may have a $\sim -10\%$ negative bias when an ensemble of measurements are considered. Notably, the increased complexity of the METEC-TR did not yield a more significant bias like that reported by Bell et al. 2017. Statistical analysis of the residuals to assess point leverage and possible outliers supports the validity of an OLS approach for the test release data. Removing the

5 largest outlier found with the Cook's test improves both OLS fits slightly to 0.97 for METEC-TR and 1.1 for CF-TR. Residual plots are included in SI Sect. 2.

95% confidence intervals (CI) for the OLS fit were calculated through bootstrapping following the method detailed in Robertson et al. 2017, whereby the linear regressions of bootstrapped data sets are calculated to assess the range of possible regressions. Bootstrapping was used because it does not require an assumption of normally distributed data (unlike the Gaussian

10 fit approach used in Sect. 3.1.1). Using this method, the OLS correlation slopes have a mean (and 95% CI) for the CF-TR of 0.96 (0.56,1.47) and 0.96 (0.66,1.28) for the METEC-TR.

3.1.3 Bland-Altman Analysis

Because the rate of methane releases for both the METEC and Christman tests are known to within a small margin of error (<5%), OLS regression, which assumes no error in the independent variable, is a reasonable approach. However, OLS analysis

- 15 is weighted by larger release rates and may not give an accurate representation of OTM 33A performance at all methane emission rates. Bland-Altman (BA) analysis removes this bias by considering the difference between the test release and OTM measurements (known release - OTM flux) as a function of a known release rate (Fig. 7) (Giavarina, 2015). Bland-Altman analysis also assumes that the method difference (y-axis) comes from a normal distribution. Kolmogorov-Smirnov statistical tests supporting the normality of the method difference can be found in SI Sect. 3. For BA analysis, if the 2σ range of method
- 20 difference includes zero, the methods are considered to be statistically equivalent; i.e. no bias (Giavarina, 2015). The BA plot also illustrates the amount by which OTM can over- (negative numbers) or under-estimate (positive numbers) the known release. On average, the CF-TR and METEC-TR both underestimate the known releases, with mean differences of 0.028 kg h⁻¹ and 0.025 kg h⁻¹ respectively. However, since the 2σ interval includes zero, the BA analysis identifies no statistical difference between the OTM 33A flux estimate and the known release rate.

25 3.1.4 Orthogonal Distance and Variance Weighted Least-Squares Regression

Other approaches for minimizing the influence of larger release rates on the OLS fit include orthogonal distance regression (ODR) and variance weighted least squares regression (VWLS). These methods take into account error in both the x and y variables, and require that each measurement has an independent uncertainty estimate on both axes. Since uncertainty of OTM 33A flux estimates is taken as a fixed percentage of the estimated value, these methods tend to perceive a higher confidence

30 (smaller absolute uncertainty) in smaller estimates, and a lower confidence (higher absolute uncertainty) in larger estimates. This results in a fit with a low bias since the estimates with smaller absolute uncertainty are strongly weighted, and less weight is given to estimates with larger uncertainties. To examine these approaches, the METEC-TR are used as an example below. Applying a measurement uncertainty of $\pm 50\%$ (representing the % error that roughly 85% of the data points are within) for each OTM measurement and the metered uncertainty for each METEC-TR in kg h⁻¹ yields an ODR slope of 0.79 ± 0.09 when the intercept is set to (0,0) (Fig. 8). A lower slope of 0.67 \pm 0.1 is found using the VWLS method. In this case the ODR and VWLS regressions suggest the OTM flux estimates are 20–33% lower than the known releases, where an OLS regression indicates the method is only 8% low. Total emissions estimated by OTM 33A (23.074 kg hr⁻¹) are 2.5% lower than the total

known emission rates (23.67 kg hr⁻¹), suggesting OLS regression is a better fit for this data set.

VWLS and ODR should be used with caution where the measurement uncertainty is not independent of the measurement (i.e. a constant fractional error), because this may discriminate against data points of a larger magnitude depending on uncertainty in the other (x or y) variable. Bell et al. (2017) used a VWLS regression to compare OTM 33A measurements to onsite

- 10 measurements of O&G production facilities in the Fayetteville, which yields yielded a correlation of 0.41 (+0.51, -0.17). The 95% CI of the VWLS regression are calculated through bootstrapping the regression and considering the uncertainties in both the study onsite estimate and the OTM 33A estimate for each data point. Repeating this analysis using an OLS regression without an intercept results in a correlation of 0.39 (+0.39, -0.15) (Fig. S6). If an OLS regression were chosen instead of the VWLS, the conclusion by Bell et al. 2017 that OTM 33A produced lower emissions estimates than the onsite measurement
- 15 results would not have been effected. The total mass flux measured by OTM 33A (13 (+5.3, -2.1) kg h⁻¹) and onsite teams (19 (+7.7, -3) kg h⁻¹) also supports the conclusion that the OTM 33A flux estimate was biased low relative to the onsite measurements at these paired facilities.

3.2 Sensitivity Analysis

5

3.2.1 OTM 33A Sensitivity to Source Distance

- 20 Because OTM 33A assumes a point source, the distance to the release point has a large influence, as this impacts the modeled plume spread and therefore the final calculated flux. The importance of an accurate source distance in Gaussian plume modeling has been noted in previous studies (Lan et al., 2015; Caulton et al., 2018). During the METEC-TR, the University of Wyoming measurement team had site access and we were able to determine the exact emission point(s) using an IR camera. With this knowledge, we were able to calculate the exact source distance, or the average distance in the case of multiple emission point(s). For sources. In the field, site access is often not available and it's often not possible to detect the most likely emission point(s). For
 - this reason, the average distance of possible emission points is used when calculating source distance.

For a 5% change in source distance, the OTM 33A flux estimate increased by almost 10% (Fig.9(a))33A sensitivity to source distance was tested two ways for the METEC test releases. The following test was performed during the data analysis stage, and compared the flux estimated using the average distance of all the components that could be sighted with the range

30 finder from the van (e.g. wellhead, separator, tank) to the flux estimated using the distance to the known release point or point distance (identified using the FLIR camera). Although the well pad measured at the METEC facility was quite small (~6 m by 6 m), the average source distance was larger than the specific source distance ~60% of the time. Source distance related error is small in the context of the ±70% measurement error, but underscores how determination of the exact emission point can

further reduce errors in the field. Looking at this change in terms of a kg hr⁻¹ flux, The change in the OTM 33A flux (Δ Flux) as a result of changing the measurement distance (Δ Distance) was found using Equations 3 and 4.

$$\% \Delta Distance = \frac{Average \ Distance - Point \ Distance}{Average \ Distance} \times 100$$
(3)

(4)

5 $\% \Delta Flux = \frac{Average \ OTM - Point \ OTM}{Average \ OTM} \times 100$

A correlation plot of % Δ Distance and % Δ Flux suggests that for a 5% change in source distance, the OTM 33A flux estimate would increase by almost 10% (Fig. 9(b)shows that a)). In terms of mass error, the OTM flux estimated by the average or specific source distance has very little impact in the over- or under-estimation of the METEC known release - (Fig. 9(b)). Allowing this fit to have an intercept changes the linear fit to y = 0.978x - 0.03, a negligible difference. Source distance

10 related error is small in the context of the \pm 70% measurement error, but this analysis underscores how determination of the exact emission point can further reduce errors in the field.

OTM 33A sensitivity to distance was also tested in the field during the METEC test releases. For configurations that had both a "closer" (generally <70 m) and "farther" (generally >100m) measurement distance for replicate measurements, the closer measurement had a flux estimate closer to the known release 78% of the time (SI Sect. 1.4). The average distance of the

15 closer replicate measurements (78 m) is comparable to the average measurement distances for the CF-TR of 78 m, smaller than the mean METEC-TR distance of 114 m, and larger than the measurement distances during the Arkansas campaign of 46 m (20–113 m) (Robertson et al., 2017; Bell et al., 2017). For both the CF-TR and the METEC-TR, there is no obvious increase in % error as measurement distance increases (Fig. 10(a)), suggesting the underestimation reported by Bell et al. cannot be blamed solely on closer measurement distances.

20 3.2.2 OTM 33A Error - Sensitivity to Wind Speed

One hypothesized reason for the underestimation of OTM 33A compared to onsite methods reported in the Bell et al. study is the lower wind speeds ($<2 \text{ m s}^{-1}$) experienced in that study. The CF-TR and METEC-TR both had wind speeds higher than 2 m s⁻¹, making an absolute conclusion impossible, but for the wind speeds measured there is no obvious trend between the mean measurement wind speed and OTM 33A error (Fig. ??)10(b)).

25 3.2.3 OTM 33A Error - Sensitivity to Number of Sources and Source Height

The METEC-TR included multiple emission points, both slightly above and below the sampling inlet height. There is no obvious trend between the number of release points and % error, though the sample size for two or more sources is relatively small (N=6). The height of the source also shows sources tested also show no obvious influence on OTM 33A accuracy.

3.3 Ensemble Mass Flux

OTM 33A measurements are often used to find a an average emission rate per well or per facility in an O&G basin (Robertson et al., 2017; Alvarez et al., 2018). To assess the accuracy of the mean of a number of OTM 33A measurements, the mean mass flux measured by OTM 33A is compared to the mean mass flux of the known release through bootstrapping. The bootstrapping

- 5 approach is used to generate more statistically robust results without the need for assuming Gaussian distributions. The OTM 33A flux estimates and known releases (including their respective measurement uncertainties) are sampled with replacement, summed, and compared following Robertson et al. (2017). This approach suggests that the addition of complexity in the METEC-TR did not significantly impact the accuracy of OTM 33A (Fig. 12), and for both sets of test releases there is a large amount of overlap between the OTM 33A and known release distributions (Fig. S6). These results also indicate OTM 33A
- 10 does not drastically underestimate the total emissions for an ensemble or group of measurements, and that scaling-up mean emissions measured with OTM 33A to an entire basin is a valid approach.

Figure 13 summarizes measurement conditions experienced during four University of Wyoming field campaigns (Robertson et al., 2017) and the test releases. Data from the four O&G basins were represented a significant fraction of data, along with other field campaigns used by Alvarez et al. (2018) to generate an estimate of national methane emissions. Measurement

15 conditions in Fayetteville, Arkansas (AR) were notable for closer source distance, lower wind speeds, and generally more unstable atmospheric conditions. All of these variables could have influenced the low bias reported by Bell et al. and are conditions that should be replicated (if possible) in future test releases.

4 Conclusions

The more realistic test releases described in this study build on preexisting test releases and suggest a single OTM 33A measurement can have a 2σ error of $\pm 70\%$. Analysis of both the simple CF-TR and more complex METEC-TR indicate that under these measurement conditions and release rates, an ensemble of OTM 33A may have a slight negative bias (~5%) when compared to a known release rate through an OLS model. The mean and 95% CI found through bootstrapping are 0.96 (0.56,1.47) and 0.96 (0.66, 1.28) for the CF-TR and METEC-TR respectively. The 40-60% underestimation reported in the Bell et al. study was not replicated during either test release experiment.

- 25 OTM 33A flux estimates are sensitive to the assumed source distance, with a +5% change in source distance corresponding to a $\sim +10\%$ change in the OTM flux. However, the error caused by uncertainty in source distance is small compared to the measurement method error determined through these test releases. During field measurements, uncertainty in source distance can be mitigated by having site access and an IR camera to detect the emission source(s). Uncertainty did not correspond to wind speeds observed during the test releases, but was relatively higher for smaller release rates. Sensitivity of OTM 33A to the number or height of amission sources was inconclusive.
- 30 the number or height of emission sources was inconclusive.

For both test release experiments, the maximum release rates $(2-2.15 \text{ kg h}^{-1})$ were constrained by available resources and facility throughput and, while they represent a large fraction of emission rates observed in the field, they do not fully encompass the dynamic range of emissions observed in an O&G basin. The bootstrapped mean emission rates from four O&G basins

measured by the University of Wyoming range from $0.68-3.7 \text{ kg h}^{-1}$ (Robertson et al., 2017), suggesting the range of these test releases may not be representative of the largest emission rates observed in the field (Fig. 13).

OTM 33A has been used to estimate mean facility emissions and basin-wide facility emissions in a number of O&G basins. The mean mass fluxes and 95% CI for each test release experiment are not statistically different. This analysis lends confidence

5 to nation wide national emission estimates from the O&G production sector using OTM 33A measurements. The Despite the OTM 33A estimated limit of detection (0.01 g s⁻¹) and relative overestimation of smaller release rates, the analyses reported here and the study by Bell et al. suggest that OTM 33A does not overestimate an ensemble of flux estimates.

Code and data availability. Available on request.

Author contributions. Authors RE, ARM, DAS, JS, SMM, and RAF collected data and helped design the study. Authors DZ, CSB, TLV
helped with study design, statistical methods, and study comparison. Authors RE, AMR, SMM, RAF, and CSB wrote and edited the manuscript. Authors DAS, JS, and DZ provided feedback on the manuscript.

Competing interests. The authors have no competing interests to report.

Acknowledgements. The authors acknowledge support from the Wyoming School of Energy Resources Center of Excellence in Air Quality, the Clean Air Task Force, and the Methane Emissions Technology Evaluation Center.

References

5

20

35

Allen, D. T.: Emissions from oil and gas operations in the United States and their air quality implications, Journal of the Air & Waste Management Association, 66, 549–575, 2016.

Alvarez, R. A., Pacala, S. W., Winebrake, J. J., Chameides, W. L., and Hamburg, S. P.: Greater focus needed on methane leakage from natural gas infrastructure. Proceedings of the National Academy of Sciences, 109, 6435–6440, 2012.

- Alvarez, R. A., Zavala-Araiza, D., Lyon, D. R., Allen, D. T., Barkley, Z. R., Brandt, A. R., Davis, K. J., Herndon, S. C., Jacob, D. J., Karion, A., Kort, E. A., Lamb, B. K., Lauvaux, T., Maasakkers, J. D., Marchese, A. J., Omara, M., Pacala, S. W., Peischl, J., Robinson, A. L., Shepson, P. B., Sweeney, C., Townsend-Small, A., Wofsy, S. C., and Hamburg, S. P.: Assessment of methane emissions from the U.S. oil and gas supply chain, Science, 2, 7204–9, 2018.
- 10 Bell, C., Vaughn, T., Zimmerle, D., Herndon, S., Yacovitch, T., Heath, G., Pétron, G., Edie, R., Field, R., Murphy, S., Robertson, A., and Soltis, J.: Comparison of methane emission estimates from multiple measurement techniques at natural gas production pads, Elementa: Science of the Anthropocene, 5, 1–14, 2017.
 - Brandt, A. R., Heath, G. A., Kort, E. A., O'Sullivan, F., Pétron, G., Jordaan, S. M., Tans, P., Wilcox, J., Gopstein, A. M., Arent, D., Wofsy, S., Brown, N. J., Bradley, R., Stucky, G. D., Eardley, D., and Harriss, R.: Methane Leaks from North American Natural Gas Systems,

```
15 Science, 343, 733–735, 2014.
```

- Brantley, H. L., Thoma, E. D., Squier, W. C., Guven, B. B., and Lyon, D.: Assessment of Methane Emissions from Oil and Gas Production Pads using Mobile Measurements, Environmental Science & Technology, 48, 14508–14515, 2014.
- Brantley, H. L., Thoma, E. D., and Eisele, A. P.: Assessment of volatile organic compound and hazardous air pollutant emissions from oil and natural gas well pads using mobile remote and on-site direct measurements, Journal of the Air & Waste Management Association, 65, 1072–1082, 2015.
- Caulton, D. R., Shepson, P. B., Santoro, R. L., Sparks, J. P., Howarth, R. W., Ingraffea, A. R., Cambaliza, M. O. L., Sweeney, C., Karion, A., Davis, K. J., Stirm, B. H., Montzka, S. A., and Miller, B. R.: Toward a better understanding and quantification of methane emissions from shale gas development., Proceedings of the National Academy of Sciences of the United States of America, 111, 6237–6242, 2014.
- Caulton, D. R., Li, Q., Bou-Zeid, E., Fitts, J. P., Golston, L. M., Pan, D., Lu, J., Lane, H. M., Buchholz, B., Guo, X., McSpiritt, J., Wendt,
 L., and Zondlo, M. A.: Quantifying uncertainties from mobile-laboratory-derived emissions of well pads using inverse Gaussian methods, Atmospheric Chemistry and Physics, 18, 15145–15168, 2018.
 - Coburn, S., Alden, C. B., Wright, R., Cossel, K., Baumann, E., Truong, G.-W., Giorgetta, F., Sweeney, C., Newbury, N. R., Prasad, K., Coddington, I., and Rieker, G. B.: Regional trace-gas source attribution using a field-deployed dual frequency comb spectrometer, Optica, 5, 320, 2018.
- 30 Conley, S., Faloona, I., Mehrotra, S., Suard, M., Lenschow, D. H., Sweeney, C., Herndon, S., Schwietzke, S., Pétron, G., Pifer, J., Kort, E. A., and Schnell, R.: Application of Gauss's theorem to quantify localized surface emissions from airborne measurements of wind and trace gases, Atmospheric Measurement Techniques, 10, 3345–3358, 2017.
 - Daube, C., Conley, S., Faloona, I. C., Arndt, C., Yacovitch, T. I., Roscioli, J. R., and Herndon, S. C.: Using the tracer flux ratio method with flight measurements to estimate dairy farm CH4 emissions in central California, Atmospheric Measurement Techniques Discussions, pp. 1–20, 2018.
 - Foster-Wittig, T. A., Thoma, E. D., and Albertson, J. D.: Estimation of point source fugitive emission rates from a single sensor time series: A conditionally-sampled Gaussian plume reconstruction, Atmospheric Environment, 115, 101–109, 2015.

Giavarina, D.: Understanding Bland Altman analysis, Biochemia Medica, 25, 141–151, 2015.

- Harriss, R., Alvarez, R. A., Lyon, D., Zavala-Araiza, D., Nelson, D., and Hamburg, S. P.: Using Multi-Scale Measurements to Improve Methane Emission Estimates from Oil and Gas Operations in the Barnett Shale Region, Texas, Environmental Science & Technology, 49, 7524–7526, 2015.
- 5 Karion, A., Sweeney, C., Pétron, G., Frost, G., Michael Hardesty, R., Kofler, J., Miller, B. R., Newberger, T., Wolter, S., Banta, R., Brewer, A., Dlugokencky, E., Lang, P., Montzka, S. A., Schnell, R., Tans, P., Trainer, M., Zamora, R., and Conley, S.: Methane emissions estimate from airborne measurements over a western United States natural gas field, Geophysical Research Letters, 40, 4393–4397, 2013.
 - Karion, A., Sweeney, C., Kort, E. A., Shepson, P. B., Brewer, A., Cambaliza, M., Conley, S. A., Davis, K., Deng, A., Hardesty, M., Herndon, S. C., Lauvaux, T., Lavoie, T., Lyon, D., Newberger, T., Pétron, G., Rella, C., Smith, M., Wolter, S., Yacovitch, T. I., and Tans, P.: Aircraft-
- 10 Based Estimate of Total Methane Emissions from the Barnett Shale Region, Environmental Science & Technology, 49, 8124–8131, 2015. Lan, X., Talbot, R., Laine, P., and Torres, A.: Characterizing Fugitive Methane Emissions in the Barnett Shale Area Using a Mobile Labora-
 - Lavoie, T. N., Shepson, P. B., Cambaliza, M. O. L., Stirm, B. H., Karion, A., Sweeney, C., Yacovitch, T. I., Herndon, S. C., Lan, X., and Lyon,
 D.: Aircraft-Based Measurements of Point Source Methane Emissions in the Barnett Shale Basin, Environmental Science & Technology,
- 15 49, 7904–7913, 2015.

30

- Lavoie, T. N., Shepson, P. B., Gore, C. A., Stirm, B. H., Kaeser, R., Wulle, B., Lyon, D., and Rudek, J.: Assessing the Methane Emissions from Natural Gas-Fired Power Plants and Oil Refineries, Environmental Science & Technology, pp. acs.est.6b05 531–9, 2017.
 MATLAP: version 8.5.0 (P2015a). The MathWorks Inc. National Massachusette. 2015.
- MATLAB: version 8.5.0 (R2015a), The MathWorks Inc., Natick, Massachusetts, 2015.

tory, Environmental Science & Technology, 49, 8139-8146, 2015.

- Omara, M., Zimmerman, N., Sullivan, M. R., Li, X., Ellis, A., Cesa, R., Subramanian, R., Presto, A. A., and Robinson, A. L.: Methane
 Emissions from Natural Gas Production Sites in the United States: Data Synthesis and National Estimate, Environmental Science & Technology, 52, 12 915–12 925, 2018.
 - Peischl, J., Ryerson, T. B., Aikin, K. C., de Gouw, J. A., Gilman, J. B., Holloway, J. S., Lerner, B. M., Nadkarni, R., Neuman, J. A., Nowak, J. B., Trainer, M., Warneke, C., and Parrish, D. D.: Quantifying atmospheric methane emissions from the Haynesville, Fayetteville, and northeastern Marcellus shale gas production regions, Journal of Geophysical Research: Atmospheres, 120, 2119–2139, 2015.
- 25 Peischl, J., Karion, A., Sweeney, C., Kort, E. A., Smith, M. L., Brandt, A. R., Yeskoo, T., Aikin, K. C., Conley, S. A., Gvakharia, A., Trainer, M., Wolter, S., and Ryerson, T. B.: Quantifying atmospheric methane emissions from oil and natural gas production in the Bakken shale region of North Dakota, Journal of Geophysical Research: Atmospheres, 121, 6101–6111, 2016.
 - Peischl, J., Eilerman, S. J., Neuman, J. A., Aikin, K. C., de Gouw, J., Gilman, J. B., Herndon, S. C., Nadkarni, R., Trainer, M., Warneke, C., and Ryerson, T. B.: Quantifying Methane and Ethane Emissions to the Atmosphere From Central and Western U.S. Oil and Natural Gas Production Regions, Journal of Geophysical Research: Atmospheres, 17, 13,941, 2018.
- Pétron, G., Karion, A., Sweeney, C., Miller, B. R., Montzka, S. A., Frost, G. J., Trainer, M., Tans, P., Andrews, A., Kofler, J., Helmig, D., Guenther, D., Dlugokencky, E., Lang, P., Newberger, T., Wolter, S., Hall, B., Novelli, P., Brewer, A., Conley, S., Hardesty, M., Banta, R., White, A., Noone, D., Wolfe, D., and Schnell, R.: A new look at methane and nonmethane hydrocarbon emissions from oil and natural gas operations in the Colorado Denver-Julesburg Basin, Journal of Geophysical Research: Atmospheres, pp. 1–17, 2014.
- 35 Ravikumar, A. P., Wang, J., and Brandt, A. R.: Are Optical Gas Imaging Technologies Effective For Methane Leak Detection?, Environmental Science & Technology, 51, 718–724, 2016.
 - Ravikumar, A. P., Wang, J., McGuire, M., Bell, C. S., Zimmerle, D., and Brandt, A. R.: "Good versus Good Enough?" Empirical Tests of Methane Leak Detection Sensitivity of a Commercial Infrared Camera, Environmental Science & Technology, 52, 2368–2374, 2018.

- Rella, C. W., Tsai, T. R., Botkin, C. G., Crosson, E. R., and Steele, D.: Measuring Emissions from Oil and Natural Gas Well Pads Using the Mobile Flux Plane Technique, Environmental Science & Technology, 49, 4742–4748, 2015.
- Robertson, A. M., Edie, R., Snare, D., Soltis, J., Field, R. A., Burkhart, M. D., Bell, C. S., Zimmerle, D., and Murphy, S. M.: Variation in Methane Emission Rates from Well Pads in Four Oil and Gas Basins with Contrasting Production Volumes and Compositions, Environmental Science & Technology, 51, 8832–8840, 2017.
- Roscioli, J. R., Yacovitch, T. I., Floerchinger, C., Mitchell, A. L., Tkacik, D. S., Subramanian, R., Martinez, D. M., Vaughn, T. L., Williams, L., Zimmerle, D., Robinson, A. L., Herndon, S. C., and Marchese, A. J.: Measurements of methane emissions from natural gas gathering facilities and processing plants: measurement methods, Atmospheric Measurement Techniques, 8, 2017–2035, 2015.

Schwietzke, S., Pétron, G., Conley, S., Pickering, C., Mielke-Maday, I., Dlugokencky, E. J., Tans, P. P., Vaughn, T., Bell, C., Zimmerle, D.,

- 10 Wolter, S., King, C. W., White, A. B., Coleman, T., Bianco, L., and Schnell, R. C.: Improved Mechanistic Understanding of Natural Gas Methane Emissions from Spatially Resolved Aircraft Measurements, Environmental Science & Technology, 51, 7286–7294, 2017.
 - Snare, D. A.: A comparison of ground-based and aircraft-based methane emission flux estimates in a western oil and natural gas production basin, M.S. Thesis, University of Wyoming, Laramie, Wyoming, United States of America, 2015.

Thoma, E. D.: Assessment of Methane and VOC Emissions from Select Upstream Oil and Gas Production Operations Using Remote Mea-

- 15 surements, Interim Report on Recent Survey Studies, Proceedings of 4th Annual Conference of the Air and Waste Management Association, pp. 1–15, 2012.
 - U.S. EPA: Other Test Method (OTM) 33 and 33A Geospatial Measurement of Air Pollution-Remote Emissions Quantification-Direct Assessment (GMAP-REQ-DA), https://www.epa.gov/emc/emc-other-test-methods, 2014.

U.S. EPA: Inventory of U.S. Greenhouse Gas Emissions and Sinks 1990 - 2017, 2019.

Sciences of the United States of America, 112, 15 597–15 602, 2015.

5

35

- 20 van Kessel, T. G., Ramachandran, M., Klein, L. J., Nair, D., Hinds, N., Hamann, H., and Sosa, N. E.: Methane Leak Detection and Localization Using Wireless Sensor Networks for Remote Oil and Gas Operations, 2018 IEEE SENSORS, pp. 1–4, 2018.
 - Vaughn, T. L., Bell, C. S., Pickering, C. K., Schwietzke, S., Heath, G. A., Pétron, G., Zimmerle, D. J., Schnell, R. C., and Nummedal, D.: Temporal variability largely explains top-down/bottom-up difference in methane emission estimates from a natural gas production region, Proceedings of the National Academy of Sciences, 121, 201805 687–6, 2018.
- 25 Yacovitch, T. I., Herndon, S. C., Pétron, G., Kofler, J., Lyon, D., Zahniser, M. S., and Kolb, C. E.: Mobile Laboratory Observations of Methane Emissions in the Barnett Shale Region, Environmental Science & Technology, 49, 7889–7895, 2015.
 - Yacovitch, T. I., Daube, C., Vaughn, T. L., Bell, C. S., Roscioli, J. R., Knighton, W. B., Nelson, D. D., Zimmerle, D., Herndon, G. P., and C, S.: Natural gas facility methane emissions: measurements by tracer flux ratio in two US natural gas producing basins, Elementa: Science of the Anthropocene, pp. 1–12, 2017.
- 30 Zaimes, G. G., Littlefield, J. A., Augustine, D. J., Cooney, G., Schwietzke, S., George, F. C., Lauderdale, T., and Skone, T. J.: Characterizing Regional Methane Emissions from Natural Gas Liquid Unloading, Environmental Science & Technology, 53, 4619–4629, 2019.
 - Zavala-Araiza, D., Lyon, D. R., Alvarez, R. A., Davis, K. J., Harriss, R., Herndon, S. C., Karion, A., Kort, E. A., Lamb, B. K., Lan, X., Marchese, A. J., Pacala, S. W., Robinson, A. L., Shepson, P. B., Sweeney, C., Talbot, R., Townsend-Small, A., Yacovitch, T. I., Zimmerle, D. J., and Hamburg, S. P.: Reconciling divergent estimates of oil and gas methane emissions., Proceedings of the National Academy of
 - Zavala-Araiza, D., Alvarez, R. A., Lyon, D. R., Allen, D. T., Marchese, A. J., Zimmerle, D. J., and Hamburg, S. P.: Super-emitters in natural gas infrastructure are caused by abnormal process conditions, Nature Communications, 8, 14012, 2017.



Figure 1. METEC facility with possible nine of the 11 release points circled. Release points include (clockwise from top of tank) tank candy cane, tank thief hatch, tank front flange, wellhead Kimray packing, wellhead hand valve packing, separator burner fuel supply, separator Kimray vent, separator PRV, and separator house PRV. Not pictured: wellhead lubricator flange and wellhead pressure gauge. The UW mobile laboratory is in the background.



Figure 2. Summed methane enhancement and total number of datapoints in each 10° wind bin (a). Average methane enhancement per 10° wind bin and Gaussian fit (b). Goodness of fit parameter R is calculated following Eq. (S1).



Figure 3. Histogram of percent error of the OTM 33A flux estimate for both Christman and METEC test releases. Data are binned in 10% error bins. Positive % error corresponds to OTM 33A over-estimating the known release rate.



Figure 4. Scatter plot of test release error and release rate. Positive % error corresponds to OTM 33A over-estimating the known release rate.



Figure 5. Box plot of METEC and Christman OTM 33A release errors binned by release rate. The rectangle contains the median value, while the edges represent the 25th and 75th percentiles. Box whiskers include the rest of the data (100% coverage). Positive % error corresponds to OTM 33A over-estimating the known release rate.



Figure 6. Correlation plot of OTM 33A-measured flux versus known release rates. Intercept is set to (0,0).



Figure 7. Bland-Altman analysis of Christman and METEC test releases. Bold black lines represent the mean difference between the test release and the OTM measurement, while the dashed yellow and red lines indicate one and two standard deviations of the method difference.



Figure 8. Correlation plot of known release and OTM 33A flux estimate for the METEC test releases. Plot includes OLS (slope = 0.92), ODR (slope = 0.79), and VWLS (slope = 0.67) regressions.



Figure 9. (a) Correlation between percent change in distance ((Avg. distance - point)/point * 100), and resulting percent change in OTM flux ((OTM avg - OTM)/OTM * 100). (b) OTM 33A flux mass error compared to the METEC known release when using average versus known source distance.



Figure 10. OTM 33A percent error compared to the measurement distance (a). OTM 33A percent error compared to the mean measurement wind speed (b). Positive % error is OTM overestimating the known release.



Figure 11. Percent error of the METEC-TR ranked by number of sources and depending on source height rank of or average height (for multiple sources). Positive error is OTM 33A over-estimating the known release rate. Height rank 1 is a ground-level emission point, height rank 5 is the highest emission point. Height rank 6 has both low, medium, and high release points. Icons indicate the size bin of the known release rate.



Figure 12. Probability Density Function (PDF) of bootstrapped mean mass emission flux for OTM measurements and the known releases. Mean and 95% CI in kg hr⁻¹ are as follows: Christman: known release - (0.54 (0.37, 0.75), OTM - (0.51 (0.34, 0.73)). METEC: known release - (0.85 (0.58, 1.13)), OTM - (0.84 (0.60, 1.11)).



Figure 13. Summary of accepted OTM 33A measurements from field deployments and test releases (right of vertical line). Basins from Robertson et al. 2017. Upper Green River Basin, Wyoming (UGRB), Uintah Basin, UT (UB), Denver-Julesburg Basin, CO (DJ), Fayetteville, Arkansas (AR). Mean statistics (from left to right) are as follows. Distance (m): 98, 114, 83, 51, 114, 78. Flux (kg h⁻¹): 2.41, 6.99, 1.51, 1.27, 0.51, 0.96. Mean wind speed (m s⁻¹): 5.3, 4.2, 3.1, 2.9, 4.1, 4.9. Stability class: 5.0, 4.9, 5.3, 3.5, 5.0, 5.4.

Contents

5	S1 OTM 33A Analysis Program S1.1 Wind and Mixing Ratio Binning S1.2 OTM 33A Data Quality Indicators S1.3 Accepted OTM 33A Flux Measurements S1.4 Replicate OTM 33A Flux measurements	S2 S2 S2 S4 S4 S4
	S2 Ordinary Least Squares Regression	S4
	S3 Percent Error - Grouped Measurements	S6
	S4 Bland-Altman Analysis	S6
10	S5 Probability Density Functions for Bootstrapped Data	S6
	S6 OLS fitting of Arkansas data	S 7

List of Figures

	S 1	Data Quality Indicators
	S 2	Replicate measurements
15	S 3	OLS residuals
	S 4	OLS leverage analysis
	S5	Combined test release - % Error
	S 6	Box plot of bootstrapped mean mass emission fluxes
	S 7	OTM33A versus Onsite flux measurements

20 List of Tables

S 1	Normality tests for BA analysis	S 6
------------	---------------------------------	------------

S1 OTM 33A Analysis Program

S1.1 Wind and Mixing Ratio Binning

The OTM 33A analysis program, written in MATLAB (2015), bins methane enhancements by 10° wind bins in order to generate a Gaussian distribution and a peak concentration. To do this, the average value of the lowest 5% of observed methane

- 5 mixing ratios are subtracted from the methane mixing ratio time series to create a series of methane enhancements above background. Then, the methane enhancements are binned by 10° wind bins to find the total enhancement for that wind direction. The number of data points that occur in each wind bin are also counted during this time. Bins that have too few measurements (or values below zero) are excluded from further analysis. The total methane enhancement per 10° wind bin (Figure 2(a)) to yield the methane enhancement per 10° wind bin (Figure 2(b)).
- 10 The R² value of the Gaussian fit is calculated as the sum of squares of the residuals divided by the sum of squares of the data following Eq. S1. Plot created using IGOR (2018).

$$R^{2} = \left(\frac{\Sigma(model \, y - mean(y))^{2}}{\Sigma(y - mean(y))^{2}}\right)$$
(S1)

S1.2 OTM 33A Data Quality Indicators

The OTM 33A analysis program includes many data quality indicators (DQI) or "flags" that help discern if an OTM 33A flux
estimate is a valid approximation. Category 1 measurements have fewer than 5 flags, Category 2 measurements have fewer than 10 flags, and Category 3 measurements have 10 or more flags. Even with these analysis flags, in-field observations of possible measurement issues and manual checks of the Gaussian fit are required.

DQI NA	ME	LIST OF EXCEEDANCE LEVELS	FLAG POINT VALUE
COUNT		< 15 minutes of measurement	3
WIND DIRECTION	ON	$(\theta_v + \theta_{vertical wind})/2 > 30$	1
		$\frac{(\theta_y - \theta_{vertical wind})}{(\theta_y + \theta_{vertical wind})/2} > 0.5$	1
TURBULENT IN	ITENSITY	TI > 0.22	1
WIND SPEED		U < 1.5 m/s	1
		$\overline{U} < 1 \text{ m/s}$	5
WIND VARIAN	CE	Wind Variance > 2.5	5
		Wind Variance > 5	10
BINNED DATA		Highest Concentration Bin $\neq 180 \pm 30^{\circ}$	1
		Highest Concentration Bin $\neq 180 \pm 60^{\circ}$	3
GAUSSIAN FIT		R < 0.95	5
		R < 0.90	10
METHANE LEV	EL	Background Methane < 1.7 ppm	1
CH4 Enhancement < 0.15 ppm & Dist. < 50 m		10	
CH4 Enhancement < 0.15 ppm & Dist. 50 – 100 m			5
CH4 Enhancement < 0.15 ppm & Dist. 100 – 150 m			3
	CH4 Enhancement < 0.15 ppm & Dist. > 150 m		
	CH4 Enhancement < 0.10 ppm & Dist. < 50 m		
CH4 Enhancement < 0.10 ppm & Dist. 50 – 100 m			5
CH4 Enhancement < 0.10 ppm & Dist. 100 – 150 m			3
CH4 Enhancement < 0.10 ppm & Dist. > 150 m			1

Figure S1. Table and caption after Snare (2015) Table 2.2.7. DQI categories and values for analysis flags. Variables include the standard deviation of the 2D wind speed Θ , mean wind speed (\overline{U}) , turbulent intensity (TI), correlation coefficient (R), source distance (Dist.) and methane enhancement.

S1.3 Accepted OTM 33A Flux Measurements

The accepted OTM 33A flux measurements and associated measurement variables are included as an Excel file.

S1.4 Replicate OTM 33A Flux measurements

During the METEC test releases, 10 of the 15 test release configurations had duplicate OTM measurements at different
distances (Fig. S2). Although this number was too small to generate robust statistics, duplicate measurements did not change the mean percent error of the OTM measurements (14.7% for replicate measurements, 13.1% for all measurements). For the configurations that had an OTM measurement that was closer to the source than the other measurement, the closer measurements generally had lower % error (seven out of nine times). Two configurations initially had more than two replicate measurements, but only the measurements with further distances passed the DQI.



Figure S2. Replicate OTM 33A measurements during the METEC test releases. The closer of the two duplicate measurements is indicated by a black square.

10 S2 Ordinary Least Squares Regression

Residuals and leverage analysis for the OLS regression are included below. The heteroskedasticity of the residuals for both fits and point leverages well below one further support the validity of an OLS model for the test release data.



Figure S3. Residuals of ordinary least square regression with intercept set to (0,0) for both test release experiments.



Figure S4. Leverage analysis of ordinary least squares regression models for CF-TR (a) and METEC-TR (b).

S3 Percent Error - Grouped Measurements

When the combined test release dataset (N=45) is fit to a Gaussian, the 1σ error is $\pm 34.5\%$. The Gaussian fit suggests a low bias of -15%. The goodness of fit parameter R is calculated using Equation S1.



Figure S5. Percent errors of both METEC-TR and CF-TR binned by 20% error bins. The 2σ error is $\pm 69\%$.

S4 Bland-Altman Analysis

5 The Bland-Altman analysis used in Section 4.3 of the main text requires that the method difference (known release - OTM 33A flux) follow a normal distribution. Normality was determined following output from a normality test package developed by Öner et al. - The results for normality are summarized below.

Test Release	Test Name	Test Statistic	p-value	Normality (1:Normal,0:Not Normal)
CF-TR	KS Limiting Form	0.8551	0.4576	1
CF-TR	KS Stephens Modification	0.8878	0.0547	1
CF-TR	KS Marsaglia Method	0.8551	0.4075	1
CF-TR	KS Lilliefors Modification	0.1866	0.0545	1
METEC-TR	KS Limiting Form	0.6719	0.7574	1
METEC-TR	KS Stephens Modification	0.6943	0.1500	1
METEC-TR	KS Marsaglia Method	0.6719	0.7071	1
METEC-TR	KS Lilliefors Modification	0.1371	0.2000	1

Table S1. Results from six normality tests for the Bland-Altman Analysis.

S5 Probability Density Functions for Bootstrapped Data

Box plots of bootstrapped PDF for the METEC and CF test releases.



Figure S6. Box plot of PDFs from Section 4.3 of bootstrapped mean mass emission flux for OTM measurements and the known tracer release. The median is marked by the red line, 25 and 75th percentiles are represented by the box edges. Whiskers extend to twice the 90th percentile, and outliers are red markers.

S6 OLS fitting of Arkansas data

Data from the Bell et al. (2017) study which compared onsite flux estimates (SOE) and OTM 33A flux estimates for the same facilities using an ordinary least squares regression in addition to the variance weighted least squares model. The VWLS (slope = 0.41) is slightly greater than that of OLS (slope = 0.39). The lower CI of both regressions overlap and are identical if rounded to 2 significant figures (0.24). The upper CI of VWLS (0.92) is closer to 1:1 than upper CI of OLS (0.78).



Figure S7. Correlation plots of OTM 33A measurements and onsite measurements (SOE) for 20 sites. (a) OLS and VLWS fits, linear scale. (b) OLS and VWLS fits, log scale (Bell et al., 2017).

References

Bell, C., Vaughn, T., Zimmerle, D., Herndon, S., Yacovitch, T., Heath, G., Pétron, G., Edie, R., Field, R., Murphy, S., Robertson, A., and Soltis, J.: Comparison of methane emission estimates from multiple measurement techniques at natural gas production pads, Elementa: Science of the Anthropocene, 5, 1–14, 2017.

5 IGOR: Version 7.08, WaveMetrics Inc., Lake Oswego, Oregon, 2018.

MATLAB: version 8.5.0 (R2015a), The MathWorks Inc., Natick, Massachusetts, 2015.

Öner, M., Applied, İ. D. K. J. o. M., and 2017: Jmasm 49: A compilation of some popular goodness of fit tests for normal distribution: Their algorithms and matlab codes (matlab), digital commons.wayne.edu.

Snare, D. A.: A comparison of ground-based and aircraft-based methane emission flux estimates in a western oil and natural gas production
 basin, M.S. Thesis, University of Wyoming, Laramie, Wyoming, United States of America, 2015.