Responses to Anonymous Referee #2

We thank the reviewer for their helpful comments which we hope will help make this work more comprehensive and easier to understand. All comments are addressed below and requested changes in text, figures, captions or tables have been included. The reviewer's comments are in bold with our specific responses below in regular text and quoted sections from the manuscript in italics.

Anonymous Referee #2 Comments

I recommend including uncertainty bounds for SS gaussian, MS gaussian and MS LES in Table 1, as well as in the abstract.

The abstract has been modified to include uncertainty ranges on p.2 line 2-3.

"The uncertainty bounds calculated for this work were 0.05q-6.5q (where q is the emission rate) for single transect sites and 0.5q-2.7q for sites with 10+ transects."

We have chosen to leave these out of Table 1 as the table shows the current state-of-the-art and our methods have yet to be explained at this point in the manuscript.

Main points:

1. Timescale

p. 3, Section 1.1: I think this section warrants a discussion of timescales of Gaussian simulations. There is a good deal of variation in what the atmospheric modeling community determines to be the appropriate timescale for the A-D stability classes. The consensus seems to be on the order of 10-15 minutes. See Fritz et al. DOI:10.13031/2013.18501 A discussion of timescales is also relevant to Reviewer #1's second comment about averaging of plume transects vs the meandering plume.

The suggested discussion has been added including the reference mentioned on p. 4 lines 20-25.

"The comparison of instantaneous and modeled concentrations is thus impacted by the averaging timescale associated with the measurements. Studies suggest that appropriate time scales depend on downwind distance and stability ranging from 2-60 min (Fritz et al. 2005). For example, in class D at 200 m downwind, 3 min would be sufficient when using the Gifford dispersion coefficients. Insufficient averaging would result in random errors that we refer to in this work as the uncertainty related to 'atmospheric variability.'"

2. Winds

p. 6, line 7: Explain why the choice was made to use downloaded interpolated hourly winds (3-hour interval data, originally) when there were presumably measured winds on-site for the intensive IGM sites? I am also concerned in how well this interpolated 1-hour met will work for determining stability parameters of quick transects. See also previous comment on timescale. How would results differ with measured met? The mobile laboratory also presumably had a measurement of mobile wind, and yet this measurement has not been mentioned at all. Do these data exist? What problems were encountered with this wind (equipment failure, uncertainty in calculating true winds for moving vehicle, etc)? Depending on the conclusions, item 2 in the Recommendations (p. 15) might be updated. Table 2 might also be updated.

Newly created Table 2 has been added to clarify when measured winds were available. The mobile laboratory wind measurements were deemed unreliable for analysis due to strong artefacts in the data. We have clarified this point on p. 7 line 31-p. 8 line 4.

"Unless measured at intensive sites with a tower, wind speed and stability are taken from NOAA's Ready Archived meteorology (https://www.ready.noaa.gov, Rolph et al., 2017) because mobile wind data showed artefacts after corrections for vehicle heading. These artefacts included unreasonably high wind speed and little correlation to stationary tower measurements."

We have added additional discussion to p. 8 lines 9-13 to clarify that the interpolation to 1-hour is indeed not ideal.

"While there is uncertainty in using interpolated model wind speed and stability, especially as conditions can change in the morning and evening, the sampling period per site lasted on timescales of a few minutes. At this temporal scale, wind speed, stability, and turbulence statistics are assumed to be constant. During rapidly changing conditions, the model interpolated wind speed and stability could indeed be incorrect. The effects of uncertainties in wind speed and stability are discussed in Sect. 5.1."

In general, this method (and all advective based emission estimates) work best under conditions that are not fast changing. The discussion of stability was intended to discuss the potential effects of using the wrong stability class. However, rapidly changing conditions are beyond the scope of this work. The mobile-mounted met station data was not used as it was deemed to be inaccurate while the vehicle was moving.

3. Clarity in describing source data

In many parts of the manuscript, it was difficult to understand the nature of the underlying dataset, or equivalently, which hierarchical "level" of data was being used. For example: p. 10, section 4.3: Are these tracer releases? Are they real emissions measured with high N and simulated with all methods?

This data scheme is based on the inverse Gaussian approach therefore we never used the tracer correlation method. However, all sites in this section were modeled with LES and had a high number of passes. We have attempted to clarify the data collection and calculation schemes used by including Table 2, which summarizes the data types and combining and simplifying figures 6 and 7 into a single figure (now Figure 6). Table S1 has also been moved to the main text as Table 3 with additional data. We have also modified some of the language throughout to better reflect the level of data being referenced. For instance, Table 4 now contains the number of transect that were released.

Fig 11: Is this purely simulated data? If measured data is available, show a sample transect on the same scale.

All the data shown in Figure 11 is simulated. This data is intended to compare the models, which simulate at a standard emission rate (1 kg/s) as necessary to compute an emission rate from Eq. 2:

$$Q = \frac{\sum C_{Observation} - C_{Background}}{\sum C_{Model}} \times Q_{ref}$$

therefore the concentrations do not match observation levels. Furthermore, we do not have real data in the vertical direction, thus we cannot show real data that matches most of these subpanels and it would

seem confusing to show the real data on only a few subpanels. The observational data corresponding to this plot are available in Figure 7.

4. Others

Table 1: Add a column to this table describing the type of uncertainty noted (e.g. 95% CI, std dev, etc. See also previous comment about adding in results of this study.

Unfortunately many authors do not explicitly explain the assumptions and considerations in their uncertainty analyses. In these cases, we assume this is meant to represent a 1 standard deviation interval as this is the most conventional metric. Additionally, some authors described their methodology, but the method is specific to their set-up and not easy to summarize in a table (e.g. Petron et al. (2012) compare a minimum and maximum range). We have emphasized further in the text that this table is meant to summarize our motivation in pursuing this work and not as an exhaustive comparison between the techniques themselves on p. 3 lines 3-11:

"These techniques have been used in different ways, from direct point source emission estimation, to area source emission rate estimation, and they use data that may span a few hours or years. For example, Kort et al. used data over 6 years at 0.33° resolution covering the entire U.S. to estimate a large emission rate. Karion et al. 2013 reported emissions from a large natural gas field (~60 km diameter) using aircraft mass-balance with data collected in a few hours while Caulton et al. 2014 reported well pad emissions (<1 km diameter) using the same technique. It is also feasible that as new instruments and data processing techniques become available, any of these techniques may be used at spatial and timescales not represented by the works cited here. Table 1 compares the author reported uncertainties for several techniques; this table is intended to illustrate the motivation for this work and not as an exhaustive review of each of these techniques."

And also regarding the uncertainty on p. 3 lines 13-17:

"Most of the author reported uncertainties appear to correspond to 1 standard deviation measurements, though, the ways these statistics are derived are not consistent. Notably, Kort et al. (2014) report a 2 standard deviation range and Petron et al. (2012) reports a minimum and maximum range for emissions. In addition, several studies report 95% confidence intervals, such as all ground-based mobile dispersion estimates, likely because of the asymmetric uncertainty that is more easily reported in this manner."

p. 7, line 1: What is the purpose of allowing the simulation to use other point sources in order to best simulate the emissions? Are these other point sources input by the user based on observed equipment/sources, or are they automatically generated? What is the reasoning behind comparing multi-point LES to a single-point gaussian simulation at all (as described p. 9, line 10 and Fig 7)? If it is for ease/quickness of data processing, describe this.

Comparing scenarios with more available data inputs is expected to reduce the uncertainty in the resulting emission rate. For instance, we normally assume a central emission source at a height of 1 m, but if we have more detailed information as to the specific location of the sources, incorporating that should reduce uncertainty in the result. We have reworked Fig. 6 to better illustrate this. More description of why the comparison was done has been added to p. 13 lines 14-16.

"A schematic of the emission rate calculation strategy is shown in Fig. 6. Generally, the more information available (e.g. source location), the less uncertain the results are likely to be. Results will be compared from different scenarios to address to what extent uncertainty can actually be reduced."

To address how sources were assigned, the readers have now been directed to Sect. 4.1 on p. 9 line 17-18 as a description seemed out of place in this section.

"A more detailed description of how sources were selected is presented in Sect. 4.1."

The description in 4.1 has also been expanded on p. 13 lines 16-19.

"As on-site access was not available and well pads may contain multiple sources, all large structures were treated as separate point sources. These were visually identified during measurements and exact coordinates were confirmed in Google Earth. The center of all sources was used as a point source. The identified sources were always gas processing units or storage tanks."

We believe the increased explanation of data availability also helps explain why the comparison is necessary as we are trying to match comparisons to scenarios with different amounts of available data.

p. 9, line 5: Did scientists have site-access or other means to verify this assumption? If tank batteries are present, tank emissions can often overwhelm wellhead emissions. Is this dealt with in the data somehow?

In this analysis we did not have site-access, see response to previous comment. We consider emissions from all sources, including tanks as the response to the previous point shows. We have added an explanation of this in Sect. 4.1. as cited in the previous response. It would be impossible to only measure only well-heads with a mobile platform at most sites thus we are only interested in emissions from the well-pad as a whole. We consider this suitable because the goal was to get integrated well-pad level emission rates. Section 5.1 discusses the uncertainty from not knowing the exact location and height of the source.

p. 13, line 9: A figure showing the distribution, where it is possible to note the mode and 95% confidence intervals, is needed to understand this skewed distribution and should be included as SI

A sample Monte Carlo distribution has been added as figure S15.

Minor comments/ typos:

Table 3. Describe SS, MS and LES abbreviations in caption.

The abbreviations in what is now Table 5 have been spelled out to avoid confusion.

Table 5: It seems like the line for Source Location should differentiate between x, y and z directions, as done in the text, or at least include a range of uncertainties.

The x and z locations have been separated in what is now Table 6. As discussed in the text, the y direction uncertainty is not expected to be significant. p. 19 line 1-2.

"Uncertainty in the cross-wind direction is not expected to contribute to significant change in emissions."

Table 6: reformat Sources of Uncertainty column (e.g. left-justify) for ease of reading.

Sources of Uncertainty column in what is now Table 8 has been left-justified.

Figure 2: Consider breaking up this graphic into more panels because it is impossible to compare individual traces as-is. Also homogenize the vertical scales. There is currently a 3 order of magnitude difference in the scales. Is this correct?

The Figure 2 caption has been clarified to explain the scale difference, which is due to the Gaussian model using a reference 1 kg/s emission rate. This is the procedure for calculating the unknown emission rate by comparing to the model with a known emission rate. To allow for the requested comparison between plume transects we have added additional subplots to what is now Figure 7 where we feel it is more useful to observe differences in the individual transects as we are talking about the actual data at this point, not describing the calculation process more generally.

Figure 4: Make the bounds of Yindex the same in all graphs. Furthermore, it would be useful to produce a mixing ratio vs Y index graph (or Scalar vs Yindex) at a height of 3m to show what hypothetical measured data would look like.

The requested changes in z and y index have been made to Figure 4.

Figure 5: Is it possible to show uncertainties at 95% confidence on these graphs? This would be more useful than the 10-90th percentiles. Replace "50% percentile" and similar with "50th percentile" throughout the caption and text

Figure 5 has been redone to improve clarity of the analysis. The original Figure 5 had the requested changes made and was moved to Figure S4.

Figure 7: expand SS, MS and LES in the graph, or note their meaning in the caption.

Figure 7 has been combined with Figure 6 in what is now Figure 6. The abbreviations have been explained.

p. 4, line 13: "Additionally, implementing..." These last two sentences seem out of place in the paragraph discussing uncertainties. Elaborate or move.

We believe describing the challenge of a controlled release is relevant to discussing how previous work has been implemented. The text has been expanded on p. 5 lines 23-24.

"As a goal of this work is to identify best practices for quantifying uncertainty, it is important to understand how feasible methods are likely to be."

p. 5, lines 1, 5, 9: LICOR is typically written as "Li-COR"

We have changed LICOR to LI-COR throughout paper.

p. 7, line 31: briefly define percent difference - what is the "true" value used as denominator.

The requested change has been made on p. 11 lines 19-20.

"The 5-95% range of observations for percent difference (pd, always relative to the mean of the compared observations)..."

p. 8, lines 2-4 and Figure 5: renumber figure panels so that consecutive letters refer to the same transect interval. As it is, line 2 should read "(b & d)" instead of "(c-d)"

Figure 5 has been redone. The original Figure 5 had the requested changes made and was moved to Figure S4.

p. 8 lines 14-19: reference Figure 6 here somewhere

Reference to Fig. 6 has been added on p. 12 line 13-14.

"As depicted in Figure 6, the goal of the sampling strategy was to produce more standard sampling sites. . ."

p. 10, line 14: which result shows no apparent bias?

Box plots of the Gaussian with on-site winds show no apparent bias. The text has been clarified to point this out on p. 16 line 27-30.

"In general, the close agreement across a range of release rates shows no apparent bias in the Gaussian with tower winds results, with the results scattered low and high relative to the release rate and no results significantly different (at the 95% CI) from the release rate."

p. 10, line 20 and Fig 10: When discussing Fig 10, the over/underestimate of results as they converge seems important to mention.

The requested discussion has been added on p. 17 lines 8-11.

"Additionally, Fig. 10 shows that the IGM rarely overestimates (>2x release rate), but more often underestimates (<0.5x release rate). By this definition the IGM results underestimated the release rate by up to 30% of the time during the release rate using only 1-5 transects."

p. 11, Section 5.1: It would be interesting to express these comparisons in relative distances as well.

Figure S12 has been expanded to include panels showing the change as the distance relative to the perturbation distance of 50 m. We believe this is the relative distance the referee was interested in, but if not please clarify and we will of course add additional figures as necessary.

p. 12, line 1: reference Table 5 in this section (Sect. 5.1).

The reference to what is now Table 6 is available in Sect. 5.2 on p. 20 lines 17-18.

"The sources for uncertainty and bias in the Gaussian measurements discussed in this analysis are summarized in Table 6."

p. 15, line 1: include range, e.g "...standard sampling uncertainty range of 0.05x - 6.0x is greater..."

The requested change has been made on p. 23 lines 20-23.

"In addition, our standard sampling uncertainty range is greater than other Gaussian approaches with controlled releases which reported 0.28q – 3.6q and 0.334q – 3.34q (Rella et al., 2015, Yacovitch et al., 2015); the uncertainty range of the multi-transect sites studies here was similar in magnitude (0.05q– 6.5q)."

Note that x has been replaced with q to reduce confusion when talking about downwind distance (also x) and emission rate.

Improving Mobile Platform Gaussian-Derived Emission Estimates Using Hierarchical Sampling and Large Eddy Simulation

Dana R. Caulton[1], Qi Li[2], Elie Bou-Zeid[1], Jessica Lu[1*], Haley M. Lane[1†], Jeffrey P. Fitts[1], Bernhard Buchholz[3], Levi M. Golston[1], Xuehui Guo[1], James McSpiritt[1], Da Pan[1], Lars Wendt[4] and Mark A. Zondlo[1]

[1]Department of Civil and Environmental Engineering, Princeton University, 59 Olden St., Princeton, NJ 08540, U.S.A.
[2]Department of Earth and Environmental Engineering, Columbia University, 500 W 120th St., New York, NY 10027, U.S.A.
[3]RMS, Technische Universität Darmstadt, Darmstadt, 64287, Germany
[4]Hunterdon Central Regional High School, Flemington, NJ 08822, U.S.A.

Correspondence to: M. A. Zondlo (mzondlo@princeton.edu)

Abstract. Mobile laboratory measurements provide information on the distribution of CH_4 emissions from point sources such as oil and gas wells, but uncertainties are poorly constrained or justified. Sources of uncertainty and bias in ground-based Gaussian derived emissions estimates from a mobile platform were analyzed in a combined field and modeling study. In a field campaign where 1009 natural gas sites in Pennsylvania were sampled, a hierarchical measurement strategy was implemented with increasing complexity. Of these sites, ~93% were sampled with an average of 2 transects in <5 min (standard sampling), ~5% were sampled with an average of 10 transects in <15 min (replicate sampling) and ~2% were sampled with an average of 20 transects ~~while~~in 15-60 min. For sites sampled with 20 transects, a tower was simultaneously ~~deploying a tower~~deployed to measure high-frequency meteorological data (intensive sampling). Five of the intensive sampling sites were modeled using large eddy simulation (LES) to reproduce CH_4 concentrations in a turbulent environment. The LES output and derived emission estimates were used to compare with the results of a standard Gaussian approach. The LES and Gaussian-derived emission rates agreed within a factor of 2~~,~~ in ~~most~~ all except one case~~s with~~; the average ~~differences of~~difference was 25%. A controlled release was also used to investigate sources of bias in either technique. The Gaussian agreed with the release rate more closely than the LES underlying the importance of inputs as sources of uncertainty for the LES. The LES was also used as a virtual experiment to ~~investigate~~determine an optimum number of repeat transects and spacing needed to produce representative statistics. Approximately 10 repeat transects spaced at least 1 min apart are required to produce statistics similar to the observed variability over the entire LES simulation period of 30 min. ~~In addition, other sources~~Sources of uncertainty ~~including~~from source location, wind speed, background concentration and atmospheric stability were also analyzed. ~~In total,~~The largest contribution to the total uncertainty was from atmospheric variability~~, observed~~; this is caused by insufficient averaging of turbulent variables in the atmosphere (also known as random errors). Atmospheric variability was quantified by repeat measurements at individual sites under relatively constant conditions~~, was found to be the most significant contributor to total uncertainty.~~. Accurate

1

~~measurements of this condition~~quantification of atmospheric variability provide a reasonable estimate of the lower bound for emission uncertainty. The uncertainty bounds calculated for this work were 0.05q-6.5q (where q is the emission rate) for single transect sites and 0.5q-2.7q for sites with 10+ transects. More transects allow a mean emission rate to be calculated with better precision. It is recommended that future mobile monitoring schemes quantify ~~this metric~~atmospheric

5 variability, and attempt to minimize it, under representative conditions to accurately estimate emission uncertainty.

1 Introduction

Reducing emissions of short-lived greenhouse gases through regulations has been considered a potentially viable way to mitigate climate change without intensively regulating CO_2, which poses economic and political challenges. In particular, reducing CH_4, a potent greenhouse gas and the main component of natural gas, may have significant immediate

10 climate change mitigation benefits (Bowerman et al., 2013, Baker et al., 2015, Zickfeld et al., 2017~~) and be a viable mitigation option.~~). However, the large numbers and types of components in the natural gas supply chain that may leak require the development of efficient and accurate methods to quantify emissions. Specifically, techniques are needed that are available to researchers at every level (government, industry and academic), accurate enough to locate and quantify specific sources of fugitive emissions and allow for self-monitoring, independent verification and understanding of common leak

15 sources. The range of emissions from these sources can be large as studies have shown that some sources, such as natural gas well pads, have a lognormal distribution where emissions span several orders of magnitude. The most critical target for mitigation efforts are the super-emitters, where the top 5% of samples can contribute ~50% of emissions (Brandt et al., 2016).

To this end, various independent CH_4 emission estimation techniques have been implemented. Table 1 shows a

20 brief summary of the methods that have been primarily applied to oil, coal and gas extraction and infrastructure which account for ~30% of the total global anthropogenic CH_4 emissions (Kirschke et al., 2013). The myriad of sites and types of emission sources have necessitated the development and application of multiple techniques. Examples include satellites (Kort et al., 2014~~)~~), remote sensing from aircraft (Kuai et al., 2016, Frankenberg et al., 2016, Thorpe et al., 2016), in-situ aircraft measurements (Karion et al., 2013 and 2014, Peischl et al., 2013 and 2015, Caulton et al., 2014, Petron et al., 2014,

25 Lavoie et al., 2015~~, Ren et al., 2017~~), long-term monitoring from short and tall towers (Petron et al., 2012), unmanned aerial vehicles (Nathan et al., 2015) and various ground-based techniques. Ground-based techniques include flask sampling (Townsend-Small et al., 2015), tracer correlation techniques (Lamb et al., 2015, Subrumanian et al., 2015, Zimmerle et al., 2015, Omara et al., 2016), chamber sampling (Allen et al., 2013 and 2014, Kang et al., 2014), thermal/optical imaging (Galfalk et al., 2015, Ravikumar et al., 2017) and combined measurement/dispersion modeling techniques from stationary

30 (Brantley et al., 2014, Foster-Witting et al., 2015) and mobile platforms (Lan et al., 2015, Rella et al., 2015, Yacovitch et al., 2015).

Every technique has ~~considerable~~various advantages and disadvantages related to operational cost, sampling efficiency, processing time and uncertainty. ~~Table 1 compares the author reported uncertainties for several techniques.~~These techniques have been used in different ways, from direct point source emission estimation, to area source emission rate estimation, and they use data that may span a few hours up to years. For example, Kort et al. (2014) used data spanning over 6 years at 0.33° resolution covering the entire U.S., to estimate a large emission rate. Karion et al. (2013) reported emissions from a large natural gas field (~60 km diameter) using aircraft mass-balance with data collected within a few hours while Caulton et al. (2014) reported individual well pad emissions (<1 km diameter) using the same technique. It is also feasible that as new instruments and data processing techniques become available, any of these techniques may be used at spatial and temporal scales not represented by the works cited here. Table 1 compares the author reported uncertainties for several techniques; this table is intended to illustrate the motivation for this work and not as an exhaustive review of each of these techniques. These uncertainty estimates should be compared with caution: self-reported uncertainties are not computed in identical manners and some may not include the same sources of uncertainty in their considerations or additional systematic biases to the same extent. Most of the author reported uncertainties appear to correspond to 1 standard deviation measurements, though, the ways these statistics are derived are not consistent. Notably, Kort et al. (2014) report a 2 standard deviation range and Petron et al. (2012) reports a minimum and maximum range for emissions. In addition, several studies report 95% confidence intervals, such as all ground-based mobile dispersion estimates, likely because of the asymmetric uncertainty that is more easily reported in this manner. The values reported for *emission* uncertainties, usually significant, are generally not from *measurement* uncertainties. Even instruments that produce high accuracy measurements that must be transformed into an emission rate can be confounded by transformation methods that rely upon limited or unrepresentative meteorological conditions such as small scale turbulence characterization, boundary layer processes or assigning background conditions to a variable atmosphere. Notable, however, is the large range of uncertainties reported by ground-based mobile dispersion techniques. These techniques rely on accurate and precise concentration measurements coupled with dispersion models (Gaussian, AERMOD, WindTrax, etc.) to produce an emission estimate and are subject to various uncertainties in the model, notably atmospheric diffusion coefficients. These techniques are attractive due to their relatively low cost, low computational requirements and high sampling efficiency of individual sources. However, as seen in Table 1, current results ~~lag behind the standard for~~using ground based and mobile measurements seem to underperform relative to other measurement techniques in this field~~.~~ in providing low uncertainty emission estimates. Therefore, there is a critical need for improved sampling methods and/or data processing in this ~~field~~area to ~~close the gap between~~improve data quality and better constrain uncertainty.

3

1.1 Theory of the Gaussian Plume Model

Approximations of scalar dispersion were investigated as early as the 1930s and were developed to describe non-reactive pollutant dispersal from elevated stacks (Sutton, 1932; Bosanquet and Pearson, 1936). As models improved, the Gaussian plume model was developed assuming that scalar concentration has a normal distribution function (Batchelor, 1949, Hilst, 1957). Additional investigation of near surface conditions where particles can either deposit to, or reflect off, the surface led to the current Gaussian plume analytical model (shown in Eq. ~~1~~1) to predict scalar concentrations (C), which can be directly derived from the advection–diffusion equation under some simplifying assumptions (as explained in Veigele and Head, 1978). Variations of this equation can be found in many papers and textbooks (for example Gifford, 1968, Zannetti, 1990). The Gaussian model with a reflective ground (CH_4 is reflected off the surface) is presented here:

$$C(x,y,z) = \frac{Q}{2\pi\sigma_y\sigma_z u} \cdot e^{\frac{-y^2}{2\sigma_y^2}} \cdot \left[e^{\frac{-(z-h)^2}{2\sigma_z^2}} + e^{\frac{-(z+h)^2}{2\sigma_z^2}} \right] \tag{1}$$

This function relies on the 3-D distances (x,y,z) of a receptor from a source as well as the source height h, mean horizontal wind speed u, and a source strength, Q. Here, the source is defined as the origin with x the downwind distance, y the crosswind distance and z the measurement height above ground. The dispersion coefficients (σ_y, σ_z) encode the strength of turbulent mixing ~~and~~or diffusivity, as well as the downwind distance over which the mixing is acted, x, which does not appear directly in Eq. 1. They are calculated according to any of several analytical parameterizations based on Pasquill-~~Giffords~~Gifford's stability class scheme. Atmospheric stability classes range from very unstable (A) to very stable (F). Class D is defined as neutral. The model describes the ~~probability density function (pdf)~~distribution of a scalar concentration (C) downwind of a source, meaning it describes average plume locations and concentrations. However, the instantaneous observed plume structure deviates greatly from the average behavior, with fluctuating peak concentration ~~location and displaying lower or higher concentration.~~magnitude and location. The comparison of instantaneous and modeled concentrations is thus impacted by the averaging timescale associated with the measurements. Studies suggest that appropriate time scales depend on downwind distance, x, and stability ranging from 2-60 min (Fritz et al. 2005). For example, in class D at 200 m downwind, 3 min would be sufficient when using the Gifford dispersion coefficients. Insufficient averaging would result in random errors that we refer to in this work as the uncertainty related to 'atmospheric variability.'

The Gaussian plume model is used to calculate emissions by comparing the model output to the observations. This can be done in a variety of ways. The stationary dispersion techniques used by Brantley et al. (2014) and Foster-Wittig et al. (2015) utilize the model at a single point and relate changes in concentration to changes in wind direction and thus speed. These procedures either follow, or are related to, the well-defined U.S. EPA OTM 33a and are not discussed further. The mobile dispersion techniques investigated in this study and others (Lan et al., 2015, Rella et al., 2015, Yacovitch et al., 2015) compare observed concentrations at continuous downwind x and y locations (i.e. along a road) to the modeled output along

this road. Various techniques have also incorporated averaging schemes and additional z dimensional data (Lan et al., 2015, Rella et al., 2015). Wind direction and speed are either fixed to the prevailing direction or rotated to match the observations. While data can be collected and processed quickly, the application of a Gaussian model that describes average plume behavior to instantaneous data has apparent shortcomings and no standard uncertainty protocol has been established.

5   **1.2 Previous Work**

Robust uncertainty analyses of Gaussian emission retrievals are not reported in most studies, which instead focus on the novel application of the methods. Yacovitch et al. (2015) reported an asymmetric 95% confidence interval on their emission rates of 0.~~334(x)~~334q – 3.~~34(x)~~34q where ~~x~~q is the reported emission rate by using a controlled release as a proxy. Lan et al. (2015) used a Monte Carlo approach based on assumed uncertainty in source height, wind speed and wind

10   direction for an average 95% confidence interval of roughly 0.~~5(x)~~5q – 1.~~5(x).~~5q. Rella et al. (2015) also used a controlled release to calculate the variation in their measurement of a constant emission and reported a 95% confidence interval of 0.~~28(x)~~28q – 3.~~6(x).~~6q. These methods report uncertainty ~~methods~~analyses that ~~are comparable in nature~~apply to a single 'site' (i.e. 1 well pad regardless of how many times it was actually sampled), though it should be noted that Rella et al. (2015) and Yacovitch et al. (2015) used downwind transects while Lan used stationary time averaged measurements.

15   Notably, Lan et al.'s (2015) Monte Carlo method produced the smallest confidence interval, but accounted only for assumed uncertainty in three parameters and may neglect other factors (distance, stability, emission variability). The controlled release method employed by Yacovitch et al. (2015) and Rella et al. (2015) is useful in that direct observations of measurement variability can be made and potential bias in the measurements can be determined. However, these methods produce large uncertainty ranges ~~that may be unsuitable to~~and it is unclear if they can reliably separate large emissions

20   (e.g. emissions >10x the mean) from ~~normal~~mean emissions. Additionally, implementing a controlled release is not trivial due to long set-up times of equipment and restricted access to locations suitable for the release. These conditions may make a controlled release experiment prohibitive for many applications with strict time or budget constraints or for those where site access is limited. As a goal of this work is to identify best practices for quantifying uncertainty, it is important to understand how feasible a given method is likely to be.

25   The Gaussian model is attractive as a method for inferring emission rates as it is fast and generalizable with the ability to account for changes in stability, wind speed and source elevation. However, the uncertainties for this method change depending upon how it is implemented and whether it is extended to situations outside the reasonable limits of the generalized form. Such situations would include using ~~an~~the average Gaussian plume ~~for unconstrained~~model with instantaneous measurements without uncertainty or ~~modeling~~sensitivity analysis or applications over complex topography.

30   A method for implementation of this technique is needed that identifies best practices and is supported by observations and modeling.

In our study, we combine traditional Gaussian methods, advanced large eddy simulation modeling and a controlled release to assess in-~~-~~situ variability of emission retrievals from CH_4 plumes downwind of natural gas well pads in the Marcellus Shale in Pennsylvania. We also investigate sources of potential bias in the controlled release and modeling methods. The basic architecture of ~~this method~~our approach to quantify errors associated with mobile Gaussian methods

5   uses (1) advanced modeling of a preselected sample site to enable investigation of optimum sampling strategies, (2) application of strategies to ~~the~~our sample collection process and (3) evaluation of additional sources of uncertainty and bias using advanced modeling and a controlled release.

2 Methods

2.1 Instrumentation

10   ~~Data~~Field data were collected in Pennsylvania during three campaigns in July 2015, November 2015 and June 2016 using the Princeton Atmospheric Chemistry Experiment (PACE). A Honda CR-V has been modified to accommodate a roof rack that ~~hold~~holds sensors ~1 m above the car, to limit the possibility of self-sampling. The roof rack is equipped with a ~~LICOR~~LI-COR 7700 to measure CH_4 and ~~LICOR~~LI-COR 7500A to measure CO_2 and H_2O; both sensors record at 10 Hz. Meteorological data and GPS data were collected at 1 Hz with a Vaisala WXT520 and Garmin unit in July 2015 and

15   June 2016 and with an Airmar WS-200WX in November 2015. More information on the mobile lab design and instrumentation can be found in Tao et al. 2015. The ~~LICOR~~LI-COR sensors were calibrated prior to each campaign using a blank (N_2) and a 2.12 ppm CH_4 standard in air and were also periodically calibrated with a 1.8724 ± 0.0030 ppm CH_4 and 394.51 ± 0.07 ppm CO_2 NOAA standard. The high stated measurement range of the LI-COR 7700 (40 ppm) and the excellent stability of the instrument allow for calibration with a relatively low concentration standard (McDermitt et al.

20   2010). In addition, enhancements observed were usually less than a few ppm above the ambient concentration. Data were synchronized and logged using a custom LabVIEW program.

In addition, at select sites, a tower was set up to measure high-frequency meteorological data. This tower included a second pair of ~~LICOR~~LI-COR 7500A and 7700 along with a METEK uSonic-3 Class A sonic anemometer to measure ~~3-D~~the three-dimensional instantaneous wind ~~components~~vector. The tower was typically set alongside the road at a height

25   between 2 and 3 m. Initially, the tower was constructed using a standard tripod, but was later adapted to the bed of a pick-up truck to allow faster deployment. The air flow around the pick-up truck was modeled (using Fluent, http://www.ansys.com/Products/Fluids/ANSYS-Fluent) to determine optimum placement of sensors above the vehicle to minimize local flow distortions. Three orientations were tested, with the truck cab facing 0°, 90° and 180° with a 0° mean wind flow. Deflection was observed in all three cases, but the distortion was minimal at ~2 m above the pickup bed. The

30   final design of the mobile lab, the instrumentation rack and the mobile tower ~~is~~are shown in Fig. 1.

2.2 ~~Site~~Hierarchical Sampling

Unconventional natural gas well pads were selected ~~before~~for sampling using a pseudo-random method to efficiently isolate sites that could be measured from public roads with the prevailing wind direction. All datasets were accessed from the Pennsylvania Spatial Data ACCESS (www.pasda.psu.edu). Sites were screened ~~for distance (<~~to remove those that were far (>300 m from public road), had obstructions (buildings and full tree lines) and large elevation difference >(>50 m~~, trees) and wind direction.~~). These characteristics were determined to be the most crucial for successfully sampling sites in this area as topography and vegetation made detecting plumes farther than 300 m difficult as the source could not be visually verified. The direction from the nearest public road was used to separate sites that could be measured with four prevailing wind conditions (N, E, W, S). This resulted in a database of screened sites for each wind direction that could be used to make measurement routes. Routes were primarily planned for efficiency around the forecast mean wind direction; however, the distribution of the sample relative to the population of key factors (well age, production, and operator) was routinely examined to identify and correct for over- and under- representation. ~~Additional~~These data were collected at different hierarchical levels based on the length of time collection took, complexity of data collected and analysis method. The data collected at each level are summarized in Table 1. A discussion of the details ~~on~~and rationale of the sampling strategy can be found in Sect. 3.

As a source of validation, an experimental controlled release of CH_4 was also performed. The controlled release allows the retrieved emissions to be compared to known emission rates from a constant source. A pure (99.5%) CH_4 cylinder was vented at various controlled flow rates to produce different measurable emission rates. The release site was selected for flat, open topography and isolated from any potential sources. Background transects were collected for approximately 30 min before the experiment to ensure no contaminating signals would be detected. An interfering signal from a large mulch pit was detected and the release set up was moved away from the source to ensure no signal mixing. The cylinder was set ~100 m from a public road at an altitude of 1 m. The release was performed over several hours, during afternoon and evening to span different stability classes. Figure S1 shows a diagram of the release set-up.

2.3 Inverse Gaussian Method (IGM)

The IGM approach has been used extensively as described in Sect. 1.1. Applied here, the method uses the sampled source location as input to first identify downwind transects. ~~Transect selection must be finalized by the user.~~ The peak CH_4 location ~~is assumed~~along with the known source location are used to ~~correspond to~~define the prevailing wind direction and centerline plume ~~direction~~ (x in Eq. 1). The along-wind and across-wind (y in Eq. 1) distances are then calculated using the synchronous GPS data. Distances are calculated for each measurement point as a transect may not necessarily be perpendicular to the wind. The receptor altitude (z in Eq. 1) is fixed at 2.5m, the height of the instrumentation above the road. Unless measured ~~on site, meteorological data including~~at intensive sites with a tower, wind speed~~,~~

7

and stability are taken from NOAA's Ready Archived meteorology ~~(https://www.ready.noaa.gov).~~(https://www.ready.noaa.gov, Rolph et al., 2017) because mobile wind data showed artefacts after corrections for vehicle heading. These artefacts included unreasonably high wind speed and little correlation to stationary tower measurements. The NOAA Ready archive meteorology dataset is from the National Center for Environmental Prediction's Eta Data Assimilation System model (EDAS, information at https://ready.arl.noaa.gov/edas40.php, Black, 1994). These climate analysis data are available in 3-hour increments ~~and are~~at 40 km resolution and constant pressure coordinates. The data were interpolated to 1 hour ~~data~~resolution for use in the model. The hourly data are matched to the closest observation based on time. The stability data are used to identify the proper z and y dispersion parameters based on Briggs (1973) for rural areas. While there is uncertainty in using interpolated model wind speed and stability, especially as conditions can change in the morning and evening, the sampling period per site lasted on timescales of a few minutes. At this temporal scale, wind speed, stability, and turbulence statistics are assumed to be constant. During rapidly changing conditions, the model interpolated wind speed and stability could indeed be incorrect. The effects of uncertainties in wind speed and stability are discussed in Sect. 5.1.

As discussed in Sect. 1.1, the comparison between the observations and modeled output along a downwind transect is used to calculate emission rate. First, the local background ($C_{Background}$), defined as the CH_4 minimum over the transect, is subtracted from observations of a plume ($C_{observation}$) to produce an enhancement value. The uncertainty of the background selection in the specific context of this work is discussed in Section 5.1. Second, Eq. 1 ~~was~~is solved for the x and y measurement points of the measured transect using a reference model emission rate (Q_{ref}) taken arbitrarily to be 1 kg s^{-1} to produce C_{model}. The ratio between the observations and model is used to infer the observed emission rate. Again, the peak observation value is used to define the plume centerline for ~~simplicity.~~the Gaussian model for each transect. A comparison of observations and model output ~~(using the reference emission)~~ from 21 downwind transects is shown in Fig. 2. Note that the roads were not necessarily perpendicular to the wind, therefore the superposition of the plume on the roadway may not show a full Gaussian profile. ~~Second~~Third, the observations and modeled concentrations ~~were~~are both integrated along y (summed since they consist of discrete points~~).~~) as previously recommended by Albertson et al. (2016) to minimize the influence of the random variability of the instantaneous plume. Finally, because the concentrations scale linearly with the emission rate according to Eq. 1, the emission rate can be estimated as shown in Eq. 2~~.~~:

$$Q = \frac{\sum C_{Observation}}{\sum C_{Model}} \frac{\sum C_{Observation} - C_{Background}}{\sum C_{Model}} \times Q_{ref}$$

(2)

~~Where~~where Σ implies summation over y. This method~~, with~~ of integrated concentrations~~,~~ has the advantage of not relying on regressions between the instantaneous data and model data which may have very low correlation as the instantaneous plume is not expected to ~~mirror a Gaussian profile on such a short time scale.~~ adhere to a Gaussian profile on such

5   comparing the observations to the Gaussian model which are described in Fritz et al. (2005).

2.4 Large-Eddy Simulation

Large-eddy simulation (LES) is used to simulate the dispersion of CH_4 for sites that had been sampled with a tower for approximately 1 hour and sampled with the mobile lab with at least 10 transects at both the beginning and end of the observation period. The LES turbulent modeling technique is the most suitable for high-Reynolds number flow and

10   dispersion in the atmospheric boundary layer. The LES code used in this study has been widely validated (Bou-Zeid, Meneveau, and Parlange, 2005; Tseng, Meneveau, and Parlange, 2006; Li et al., 2016). Briefly, the LES code solves the resolved continuity, Navier-Stokes, and scalar conservation equations on a Cartesian grid, and models the unresolved motions using the Lagrangian scale-dependent dynamic subgrid-scale model (Bou-Zeid, Meneveau, and Parlange, 2005). The sharp interface (Mittal and Iaccarino, 2005) immersed boundary method is used to simulate flow with the presence of

15   large solid structures (e.g. tanks in this study) in the field (Chester, Meneveau, and Parlange, 2007; Li, Bou-Zeid, and Anderson, 2016; Tseng, Meneveau, and Parlange, 2006). ~~Volumetric scalar~~Scalar sources are located on top of the structures or at other points around the source structure if needed to simulate the gas emissions. A more detailed description of how sources were selected is presented in Sect. 4.1.

A pseudo-spectral method is used for horizontal spatial derivatives and a second-order finite difference method is

20   used for vertical spatial derivative with the needed treatments to overcome the Gibbs phenomenon following Li, Bou-Zeid and Anderson (2016). Second-order Adams-Bashforth method is used for time integration. The inflow velocity is a turbulent logarithmic profile generated from a separate simulation over homogeneous flat terrain. mimicking upwind conditions. The inflow scalar is kept at a constant background concentration.

In total, 5 sites were simulated. in neutral conditions. Most sites were set up with 1 or 2 m horizontal and vertical

25   grid resolution with total simulation domain size of 256 m in x (along-wind) and y (cross-wind) directions and 100 m in z (vertical) direction. Site 5, the controlled release, was set up with 1 m horizontal resolution and 0.2222 m vertical resolution with a full z dimension of 33.33 m~~; this was done to improve resolution of a source at low elevation.~~. This was done to because the release source is at a low elevation and such high resolutions are needed by LES to resolve a sufficient fraction of the turbulent scales near the surface. It is important to note that all sites were set up to ensure that the vertical

30   dimension was at least 10× the tallest simulated obstruction, which is equal to the height of the emission sources. This is significant as according to Townsend's theory of attached eddies, turbulent scales that contribute significantly to vertical diffusion of a plume are proportional to the distance of this plume from the surface (thus 'attached' to the surface, Townsend

1961). Townsend's theory has been confirmed experimentally by many studies (Perry and Li 1990, Nickels et al. 2007, Woodcock and Marusic 2015) for near-neutral conditions. Since the domain size of the LES will limit itself the largest scale of eddies that can be resolved, the large z dimension used for these sites will allow full contributions from eddies of size up to one order large than the emission point elevation, which will diffuse and spread the plume. The real atmospheric boundary

5　layer might contain even larger eddies (up to ~1 km) that are not captured in the LES; since these eddies are much larger than the source height and thus the plume cross sectional scale, they will cause plume meandering, but will not diffuse the plume in the y and z directions. As a result, LES might underestimate plume meandering, a point that will be revisited in Sect. 4.2.

　　　　　Site layouts are shown in Fig. 3. Sites were simulated for at least 30 min to allow the simulated turbulence to reach

10　~~steady~~a statistically stationary state~~.~~, where average and standard deviations of wind and scalars approach a constant value as shown in Fig. S2. The equations solved in LES ~~output~~are non-dimensionalized using the friction velocity (u_*). The advantage is ~~rescaled~~that results from LES apply more broadly to any problem when the non-dimensional quantities, e.g. as $u_{nondimensional} = u / u_*$, are considered. LES outputs can then be scaled (dimensionalized) with the measured field friction velocity and the scalar flux rate imposed in the simulation ~~(i.e. to get a LES~~ to match the reference emission of Q_{ref} = 1 kg s^{-1} ~~) to~~

15　~~allow~~ that allows direct comparison to the Gaussian model estimates. Table ~~S1~~2 summarizes conditions and domain parameters for all 5 sites. Sites were primarily selected for simple geometry with flat terrain and homogeneous upwind conditions. Generally, elevation differences across the domains were less than 4 m and structures could be easily seen and photographed from the road to aid in site set-up.

3 Sampling Strategy

20　### 3.1 Model-Based Design of Sampling Strategy

　　　　　~~As the~~ LES ~~output~~has been previously used to investigate plume dispersion and is used here as a reference that represents the best estimate for the 'truth' of how a plume evolves in a turbulent near-neutral environment~~, a~~ (Nieuwstadt and de Valk, 1987, Weil 1990. Wyngaard and Weil, 1991, Mason 1992, Weil et al. 2004). A useful extension of the LES analysis would be to examine the output as a reference case to understand how 'sampling' the model environment by taking

25　instantaneous 'measurements' of the concentration fields affects emission retrievals. The turbulent structures that LES can resolve are illustrated in Fig. 4, which contrasts instantaneous plumes and averaged plumes in both the horizontal stream-wise (x-y) and vertical cross-wind (y-z) perspective. To optimize sampling there are two important variables (1) the number of measurements and (2) the time interval between measurements. Increasing the number of measurements is expected to increase the accuracy of the retrieval; however, the time interval may also affect results as measurements with short spacing

may ~~sample similar coherent plume structures.~~resample the same coherent plume and thus the same plume realization (Metzger at al., 2007, Shah and Bou-Zeid, 2014).

Using the LES output, ~~random samples were picked~~ which is saved with 1 Hz resolution to match our instrument sampling frequency, sample transects were picked from ~~1-N samples (N being~~ the ~~total~~ full time series, ~~varying the~~ number of ~~available samples for a given scenario) with~~ repeat transects and their time intervals. Time intervals of 30s, 1 min, 2 min and random~~-~~ (meaning the time interval was not consistent or constrained) were imposed upon the sample picks and the number of repeat transects ranged from 1 to 70. These 'transects' were then integrated ~~to replace~~and used as $C_{observation}$ in Eq. ~~2~~2 (LES is the experiment here) and the average LES profile was used as C_{Model} to produce an emission rate. ~~As~~For each combination of transect number and time spacing, 100 random samples were picked and the mean and standard deviation of the emission rate were calculated and compared to the known LES emission rate. An ideal scenario would result in a low percent difference and a standard deviation of the sample that is roughly equal to the standard deviation of a fully random sample (random time spacing). The random time spacing should be representative of a fully random sample as points are drawn from the full 30 min time simulation and are less likely to resample similar plume structures. Standard deviations are being compared instead of standard error as each sample strategy is being treated as a population so that the resulting standard deviation may be used as an approximation of the population standard deviation. Box plot of the results for the 100 random samples are shown in Fig. 5 ~~(a-h) the~~. The effect of increasing the number of measurements clearly reduces the range of retrievals, but the benefits of adding transects ~~slow down~~becomes incremental around 10 samples beyond which increasing the number of transects reduces the retrieval scatter very slowly~~.~~ (individual box plots are available in Fig. S4). The 5-95% range of observations for absolute percent difference (~~pd~~always relative to the mean of the compared observations) decreases by 60% at 10 transects, but only decreases by an additional 10% by extending up to 70 transects. The 5-95% range of observation for relative standard deviation (rsd) follows a similar but less extreme pattern with the range of observations decreasing by 20% up to 10 transects and decreasing by an additional 15% by extending up to 70 transects. Additionally, retrievals with 30s spacing ~~(c-d)~~ show increased bias (seen by higher absolute pd) as even high numbers of samples may measure plume structures that are similar as indicated by the low scatter, but are not very representative of the whole simulation. However, the 1 min ~~(e-f),~~, 2 min ~~(g-h)~~ and random ~~(a-b)~~ intervals look very similar indicating a 1 min interval can be used as a practical lower limit (this might somewhat depend on turbulence intensity and stability in the atmosphere however). Notably, the random time spacing sample shows an rsd of ~25% even at the maximum number of repeat measurements (~~N~~70). This confirms that there is ~~variability~~variation in the quantified emission rate expected simply due to atmospheric variability. Atmospheric variability may have many sources; in these ~~simulation~~simulations, variability is ~~attributed~~only attributable to turbulence. However, in a real dynamic environment atmospheric variability could also include effects of mean wind flow change and plume meandering, especially in low wind

11

speed conditions (Vickers et al., 2008, Mortarini et al., 2016). ~~These results indicate that in order to sample such that the measurements reflect the actual variability in the atmosphere~~Another way of describing the atmospheric variability as used in this work would be transect to transect variability, which encompasses all random errors that lead to differences between one transect through the plume and the next. These are hence the errors associated with insufficient averaging of the turbulent field and would be reduced as the number of averaged transect increases (Salesky and Chamecki, 2012). These results indicate that in order to sample such that the measurements minimize the effect of these random error, sites must be sampled with at least 10 transects with >1 min spacing.

3.2 Field Implementation

Field measurements were designed to target neutral stability found in the morning and evening with each sampling outing typically lasting four hours~~.~~; this minimizes the errors related to assigning stability and coincides with LES conditions. Most sites were sampled 1-3 times (denoted as standard sampling), occasional sites were sampled with ~10 transects (replicate sampling) and a few sites were sampled with >10 transects as well as a tower (intensive sampling). Typically 2 replicate sampling sites were picked per outing to capture atmospheric variability for a given condition. ~~The~~ As depicted in Figure 6, the goal of the sampling strategy was to produce ~~1000~~more standard sampling sites, ~~100~~with fewer replicate sampling sites and ~~10~~even fewer intensive sampling sites. This was based upon the approximate amount of time to acquire each sample and the limited amount of time to collect samples overall.

Field campaigns were deployed in the Marcellus shale spanning northeast and southwest Pennsylvania. In ~~the end~~total, 940 well pads were sampled with standard sampling, 53 with replicate sampling and ~~17~~16 sites with intensive sampling. These replicate sampling sites were generally chosen at the beginning and end of each four hour sampling period to observe changes in variability over the course of the sampling period that may be due to changes in atmospheric conditions. For the population of standard sampling sites with multiple passes, the average rsd of ~~emissions~~ repeat passes was 67% and the average maximum percent difference between emission estimates at a single site was 58%. The average rsd of the population of 53 emissions estimates for the replicate sampling sites was 77% and the average maximum percent difference (highest observational deviation from the mean of repeat measurements) was 150%. The rsd ranged from 12% to 260%. These populations offer insight into how sampling strategy may change estimates of these statistics and offer the chance to compare real results to the LES results shown in Sect. 3.1. These results are consistent with the LES results shown in Sect. 3.1 predicting small numbers of transects will yield an artificially low rsd and more transects are needed to produce an accurate measure of variability. Additionally, the lower maximum percent difference for standard sampling is consistent with the Sect. 3.1 LES results showing few transects will sample more similar plume structures. While there is a large range in the rsd observed, ~75% had rsd values less than 100%.

4 Source Strength Determination

4.1 Source Strength Determination Strategy

Comparisons between the IGM calculated emission rates and LES output should be done with care because the LES cannot be scaled to different distances and wind angles easily. ~~In addition, the~~The base scenario for the Gaussian approach at all ~~standard~~ sites assumes there is only one source at the 1 m elevation well-head location because the well-head is the only geolocated structure in a public database and is the only structure common at every site. Sites may have varying numbers of well-heads, but they are generally very close together (<10 m) so a centralized point is used for sites with multiple well--heads. However, the Gaussian can in fact be adjusted to include multiple sources and source heights. In order to ensure that the differences between outputs are due to the calculated model diffusion and not differences in model set-up, we compare three scenarios: (i) the base scenario is the IGM approach used for all sites that assumes there is a single source at the well-head at 1 m (SS Gaussian), (ii) the second scenario assumes the sources are other structures on the domain (i.e. storage tanks and processing equipment) that are taller and uses a multi-source Gaussian (MS Gaussian) model where all sources have the same strength and (iii) the LES that simulates sources at the same locations and heights as the MS Gaussian. A schematic of the ~~source determination strategy is shown in Fig. 7~~emission rate calculation strategy is shown in Fig. 6. Generally, the more information available (e.g. source location), the less uncertain the results are likely to be. Results will be compared from different scenarios to address to what extent uncertainty can actually be reduced. As on-site access was not available and well pads may contain multiple sources, all large structures were treated as separate point sources. These were visually identified during measurements and exact coordinates were confirmed in Google Earth. The center of all sources was used as a point source. The identified sources were always gas processing units or storage tanks.

To compare to the LES results, the observations were indexed to coordinates on the 256 by 256 m horizontal LES grid. The resulting transects were interpolated within the range of observations to account for grid cells with multiple data points or missing data points. The LES time series spanning ~30 minutes were averaged to produce a pseudo Gaussian distribution excluding a ~5 min ~~warm up time.~~initialization period (time until stationary state is achieved). LES statistics (mean and std. dev. of scalar and wind components) were plotted as a function of time to determine the onset of a steady state (Fig. S2). Because all LES runs are non-dimensionalized, the LES output needs to be scaled to represent the actual field conditions for a given observation. The LES ~~output~~ non-dimensional and dimensionalized (denoted with a superscript D) outputs can be related by Eq. 3, which can be inferred from the stationary advection diffusion equation or by analogy to the scaling of the Gaussian model in Eq. 1:

$$\frac{C_{LES}}{\dot{M}_{LES}} \times u_{LES} = \frac{C_{LES}^{D}}{\dot{M}_{LES}^{D}} \times u_{LES}^{D} \tag{3}$$

Rearranging Eq.3, the LES non-dimensional scalar output (C_{LES}) was ~~scaled~~ dimensionalized (C_{LES}^{D}) to units of kg m^{-3} according to Eq. ~~3 where M~~4:

13

$$C_{LES}^D = \frac{\dot{M}_{LES}^D}{\dot{M}_{LES}} \times \frac{u_{LES}}{u_{LES}^D} \times C_{LES} \quad\quad\quad\quad\quad (4)$$

Here, $\dot{M}_{LES}^D$ is the normalized mass (1 kg s^{-1}) and $\dot{M}_{LES}$ is the actual emission rate introduced into the ~~simulation and u$_*$ is the tower-observed friction velocity. For some sites the LES generated winds did not~~LES system. Normalizing by $\dot{M}_{LES}$ is necessary to match ~~observations (due to various potential input error sources such as surface roughness), so an additional correction factor was applied to the retrieved emissions as shown in Eq. 4.~~the reference emission rate (Q_{ref}) also used in the Gaussian retrievals. The ~~prime indicates corrected values and the corrected~~non-dimensional wind speed ~~comes from averaged tower observations.~~(u_{LES}) is divided by the dimensionalized (u_{LES}^D) wind speed, which is set equal to the on-site tower measurements. The peaks of the interpolated filed observations were centered to align with the peak location of the LES average plume as shown in Fig. ~~8;~~7; as previously discussed LES does not replicate the small changes in wind direction that can occur in the real-world that cause meandering (unless they are known and imposed).~~-~~ The LES scaled concentration at 3 m (the mobile lab measurement height) was treated as C_{Model} in Eq. 2 to produce LES derived emission rates. This methodology is used to calculate the LES emissions shown in Sects. 4.~~2~~3 and 4.~~3~~4.

$$C_{LES}^S = \frac{C_{LES}}{M\,u_*} \quad\quad\quad\quad\quad\quad\quad\quad (3)$$

$$Q_{LES}^\prime = \frac{u_{LES}^\prime}{u_{LES}} Q_{LES}$$

$$(4)$$

4.2 Clarification of 'Meandering' and a Conceptual Proof of Plume Centering

As terminology in boundary layer meteorology can be ambiguous and is not always used consistently, the specific meaning of 'meandering' is discussed here. Many previous boundary layer works have used 'meandering' to describe the general movement of a plume that does not correspond to the time averaged profile, meaning it includes all scales of motion (see Venkatram and Wyngaard 1988). However, Gaussian models do not simulate large scale meandering or shifting wind directions unless these are included in the diffusion coefficients. For instance, Seinfeld and Pandis (1998) acknowledge that plumes diffuse more with increased averaging time, which is in line with the previously discussed Fritz et al. (2005) work that observed there are optimum time scales to match a plume horizontal dispersion profile to the modeled output and observations can be quite different if their timescales are longer or shorter. The larger scale motions would be responsible for creating a broader mean plume, but do not diffuse the instantaneous plume cross-section (since they are larger than the plume's diameter). As discussed earlier, the size of eddies important for vertical diffusion are on the order of the source height; consequently, we also assume that larger scale motions do not affect the vertical profile. Deardorff and Willis (1988) suggested that the downwind plane could be redefined every 20 min to remove larger mesoscale and synoptic scale motions.

Thus we clarify that 'meandering' as used in this work corresponds to the effect of larger (>> plume diameter) scale motions (>~100 m).

As a conceptual exploration of the effects of the centering of the Gaussian on individual plumes (as described in Sect. 2.3), a simulated meandering plume with a 100 m period and 10° wind shift was investigated. The simulated plume and a comparison of a single Gaussian plume are shown in Fig. 8 (panels a and b). Using a 100 m downwind transect, the reconstructed plume, created by averaging the re-centered individual components, is compared to the Gaussian with an average wind direction and the average of all the meandering profiles. As shown in Fig. 8 (panels c and d), the exact horizontal plume structure is duplicated simply by re-centering and averaging the individual plume realizations and the vertical plume structure is always preserved whether additional meandering is included or not. When calculating emission rates using the IGM approach in this work, the instantaneous Gaussian plumes centered to the observations are used, instead of using a single average Gaussian profile. However, while the individually calculated emission rates can vary when using the aligned or average of the aligned Gaussians, the differences are expected to be small.

To investigate whether the difference between the approaches would lead to a significant bias, the results for the controlled release experiments were compared for three approaches, the Gaussian centered on each observation that we have employed, the average of the individually simulated Gaussian profiles for multiple observations, and a single Gaussian that matches the apparent average wind direction (Fig. S3). The results show that all three scenarios are virtually identical. Additionally, the centering of the Gaussian on the plume is expected to be more suitable as suggested by the standard deviations that are 0-20% lower than results that rely on a single average Gaussian. This is likely due to the occurrence of wind conditions that are not symmetrical about the centerline and road geometries that allow for plumes of closer and farther distances. This 'centering' can be thought of as a correction for the apparent conditions not matching ideal conditions in which a single Gaussian in the mean wind direction is theoretically appropriate. Yacovitch et al. (2015) also employed Gaussian centering to observations on the assumption that the source location was better constrained than the wind direction. Foster-Wittig et al. (2015) explored a similar concepts of applying corrections to their methodology, which is based on the EPA OTM33A protocol, to account for conditions that did not match the assumptions of the EPA protocol. They also observed that the differences between their methods were small. Overall, we expect uncertainty from the methodology we have employed to be constrained by the analysis present here, including using Gaussian and LES outputs that are not varied based on the observed peak location. It should be noted that this centering may not always be necessary or optimal based on the specific goal of a campaign that might include source localization or atmospheric dispersion quantification.

When centering the observations to the LES (which cannot be modulated to match apparent changing wind direction in the way the Gaussian can) the sum of the horizontal (y) distribution is conserved whether the observations are centered or not. This means that the emission rate does not change whether centering or not centering the observations. The only potential for deviation is caused by moving part of the observations outside the imposed 256 m window during centering. The plume centering that we have chosen to apply to the observations is, in this case, expected to correct for the large scale plume meandering not simulated in the LES and will not affect the calculated results; it is a way to filter out the

15

meandering impact of the large scales in the field observations. Filtering observations for the purpose of matching the domain on the LES is also discussed by Agee and Gluhovsky (1999ab) and Horst et al. (2004). Agee and Gluhovsky (1999ab) specifically address the need to remove the influence of larger scales from observations to appropriately compare observations to the LES results. Finally, it is important to highlight that the integrated plume is used to calculate emissions, which mitigates the potential for concentration mismatch when comparing instantaneous plumes with structure to averages LES or Gaussian output due to averaging timescale differences. As noted in Albertson et al. (2016), the use of an integrated plume removes the influence of the random nature of the instantaneous plume, making this a beneficial procedure for these kinds of measurements.

4.3 Controlled Release

The controlled release experiment (Site 5) utilized 3 leak rates: 0.97 $\pm$ 0.01 kg hr^{-1}, 0.22216 $\pm$ 0.002 kg hr^{-1} and 0.09090 $\pm$ 0.002 kg hr^{-1}. Release rates were controlled by two MKS mass flow controllers with stated flow accuracy of 1% of the set point (Model GE250A) or 0.05 SLM (Model 1179C). Site set-up was discussed in Sect. 2.2 and specific site details are available in Table 3. Due to the low height of the release, the LES domain was modified to have vertical resolution of 0.2222 m with a total domain height of 33.33m. Boxplots of the populations of emission retrievals from three scenarios are shown in Fig. 9 and statistics are summarized in Table 2.4. Figure 9a shows the emission rates using NOAA winds while Fig. 9b shows the emission rates calculated with measured wind. The two Gaussian retrievals are shown to compare using NOAA winds, which is the base scenario for all sites, and using the in-situ measured wind. Since the controlled release used one release point, there is only one source at a known height and the SS Gaussian approach is used. The agreement increases greatly when using the in-situ wind data as; NOAA overestimated the winds during the latter two release rates.

The Gaussian approach with in-situ measured wind agrees quite well with the release, surprisingly better than the LES. This may be due to effects of stability the Gaussian can account for, but were not simulated in the LES where neutral conditions were assumed; this will be further discussed in Sect. 5.1. In this case, the conditions shifted from slightly unstable to neutral during the second release, with the friction velocity (an important scaling parameter for LES) decreasing from 0.21 m s^{-1} to 0.10 m s^{-1}. The slightly better performance of the Gaussian method in the experiment suggests that the correction factors for stability are at least as important an input as the direct calculation of diffusion in LES across a dynamic environment. While the LES investigation is useful to investigateassess sources of error induced by sampling strategy, a controlled release is the most direct way to detect sources of bias, which otherwise would not be apparent. In general, the close agreement across a range of release rates shows no apparent bias in the Gaussian with tower winds results, with the results scattered low and high relative to the release rate. and no results significantly different (at the 95% CI) from the release rate. The Gaussian with NOAA winds showed no bias when the winds matched with observation. While LES can readily account for unstable or stable conditions, this would come at a largean increased computational cost as multiple

demanding simulations would be needed. The results here show that this ~~is~~ in fact may not be necessary, at least over flat homogeneous terrain, as the Gaussian model provides a comparable performance at a very small fraction of the modeling effort.

The controlled release was also used as an observational constraint to investigate the sampling strategy identified by the LES. This was done by randomly selecting an increasing number of transects from each release and comparing the ~~averages.~~inferred release rate using the IGM. The results in Fig. 10 are in excellent agreement with the LES results pattern seen in Fig. 5 where the average converges beyond 10 transects. This reiterates the importance of the sampling protocol and also shows the range of results possible if only a limited amount of transects are used. ~~During each release~~Additionally, Fig. 10 shows that the IGM rarely overestimates (>2x average release rate), but more often underestimates (<0.5x average release rate). By this definition the IGM results underestimated the release rate by up to 30% of the time during the release rate using only 1-5 transects. During release 1 and 3, even though a constant source is being emitted, a few (1-3) transects showed no observed plume whatsoever.

4.~~3~~4 Intensive Field Sites

Comparison between ~~means~~mean emission rates calculated from all available transects for sites 1-4 are shown in Table 5 and site specific details are shown in Table 3. Relatively good agreement is found between the LES emission retrieval and the SS and MS Gaussian approach for Sites 1-4. The range of emission retrievals vary within a factor of two with average differences of 44%. Due to the effort to standardize the comparison between all approaches by centering and correcting the observations, the difference in emission is entirely due to the difference in the dispersion each model produces. The Gaussian models assume stability corrected diffusion coefficients from Briggs (1973) while the LES makes no assumptions and allows turbulence to be numerically solved. The LES, however, was only run under neutral stability in this study. As shown in Fig. 11, the horizontal dispersion generally matches well between the LES and the MS Gaussian, while the vertical dispersion exhibits slightly different behavior. In the sites studied the LES predicts peak vertical flux (obtained by multiplying mean concentration × mean wind speed) at lower altitudes closer to the source and at higher altitudes farther from the sources with the equivalence point around 100 m downwind (additional figures of Gaussian and LES dispersion are shown is Figures ~~S3~~S5-S7). The differences are due to the distance scaling in the Briggs (1973) model being different from the LES. While there are other analytical models for distance specific dispersion coefficients, the Briggs 1973 actually matches the LES profiles better across the range of sites explored here than several other common models (Gifford, 1976, Smith, 1968; see Figures ~~S6-S9).~~ S8-S11 and Appendix A). Vertical dispersion is generally more important than the horizontal dispersion as integrating across a transect will effectively nullify any differences resulting from horizontal diffusion. However, if a difference in vertical dispersion exists, this can ~~drastically~~significantly change the retrieved emission rate. Without observations at multiple heights, it is impossible to verify which assumption is correct. In

17

the range of distances investigated in this study (<200 m) the overall discrepancy between the different model outputs is, however, small.

5 Uncertainty Analysis and Discussion

5.1 Other Uncertainty Sources

For this analysis, we have assumed the source to be constant during the time span of the measurements (typically less than 1 hour). This may not be true for all sites and may be a driver of variability. ~~Regardless, we assume that source emission variability at this scale should be treated as measurement variability as~~, for instance tanks are known to emit sporadically and Goetz et al. (2015) have shown emissions varying over the course of a few hours. However, it is not clear that there is a ~~reason~~need to quantify emission variability at scales less than 1 hour. ~~.~~ for most sources as there is a practical limit to the time resolution that can be included in inventory estimates, for instance. We thus expect any changes in emission rate at <1 hour to be reflected in what we have termed atmospheric variability, or transect to transect variability.

Other sources of uncertainty ~~we have investigated and found to be negligible include~~considered are source location and source height ~~in most cases~~. While well pads can be a few thousand m^2 in area, infrastructure that could generate leaks is usually clustered such that observed potential sources span a range of 50 m. ~~Source locations were changed for Sites 1 and 2 as shown in Table 4 according to the location of potential sources including a wellhead (1), a gas processing unit (2) or a storage tank (3) and the resulting emission retrieval were compared. Changing the across-wind location has virtually no effect on emission retrieval, while changing the along-wind location can potentially change the emissions.~~ This can be investigated theoretically by comparing the expected model sum as a function of distance assuming a 50 m shift in source location as shown in Fig. ~~S10~~S12. This scenario assumes typical conditions observed in this dataset (3 m receptor height, 1.5 m s^{-1} wind and neutral stability). Generally, changing the along-wind location of the sources changes the emission retrieval by less than 35% when measuring at >100 m downwind. However, at closer distances where the uncertainty in source location is on the order of the downwind distance this could be a major source of error. For reference, the median distance between observation and sources in this dataset is about 200 m with no sites closer than 30 m and only 5 sites less than 50 m. At 200 m uncertainty is expected to be ~20%. We also investigated the sensitivity of source height, which we estimate ranges from 1 m for wellheads to 8 m for some large storage tanks, as shown in Fig. ~~S11~~S13. The results indicate that source height variation changes the emission retrieval by less than 15~~%,~~% at downwind distances ~200 m, again due to

18

the large distance from the source. Uncertainty of the source location in the cross-wind direction is not expected to contribute to significant change in emissions.

Additionally, as shown in Sect. 4.1, inaccurate wind data can be a potential source of error. Because the modeled CH_4 concentration scales with wind in both the Gaussian and LES models, uncertainty in this parameter is necessary to constrain. In the context of this analysis, we compared the NOAA wind to the mean on-site tower measurements of winds at ~~18~~16 tower sites. NOAA wind speeds reported higher and lower values than the mean tower winds; the absolute difference was ~~differed from the tower data~~ 50% on average ~~by 50%~~. Given the linear relationship between the inverse of wind speed and Q in the Gaussian equation (Eq. 1), relative uncertainty in the wind speed should produce the same magnitude relative uncertainty in the emission rate.

An important consideration is the assignment of the background value to calculate plume enhancement. For this work, the background was calculated as the minimum value from the plume transect because the averaged 1 Hz data generally showed a uniform background near the plumes. This criteria was also compared to a scenario where the background was calculated from the average of the lowest 2% of observations in a transect with very similar results again indicating that the background value is very stable. Across all ~1000 samples the background had a median standard error of 5 ppb. When the 5 ppb tolerance is applied to the same data set, the median change in calculated flux was 4.4%. This means that the background is expected to contribute an additional ~5% uncertainty. It should be noted, however, that for very low signals the background can become a major source of uncertainty. The average peak enhancement was ~1250 ppb with a median value of ~260 ppb. Signals in this data set were screened to remove sites with peaks less than 50 ppb as this was deemed to be below the limit of detection of our system.

The final additional source of uncertainty investigated pertains to stability. The stability class determines the analytical equation used to derive the diffusion coefficients, thus affecting the emission rate. By again comparing a theoretical case, the effect of changing the stability class ~~can be seen in Fig. S12.~~by 1 can be seen in Fig. S14. The tolerance of 1 stability class reflects the fact that the atmosphere does not usually change multiple stability classes rapidly at a scale of 1-3 hours (i.e. a change from class A to class E would not be feasible). While there are certainly cases where the atmosphere can change rapidly in this time period (i.e. from class B to D), generally a miscategorization class difference of at most 1 is expected. For instance, neutral stability was targeted (Class D) for the ~1000 sites measured as part of this work. Class D was the most frequent stability class observed, with 90% of the data occurring between +/- 1 stability class. Making the stability class less stable will decrease the modeled concentration and consequently increase the emission retrieval while making the stability class more stable will have the opposite effect. The magnitude of the difference between ~~consecutive~~+/-1 stability ~~classes~~class is relatively consistent at farther downwind distance, averaging a change of 40~~%.~~% at 200 m downwind.

Not investigated here, but potentially very important to uncertainty, is the effect of terrain including both non-uniform slopes and structures such as trees. In this analysis, we have intentionally sampled sites that were determined to be

19

relatively flat and open. All of the sites modelled in this study follow this criteria, even though not every site in our sample is as simple. The geometric mean of the absolute terrain slope, defined as the absolute value of terrain rise over the distance between sources and observation, for all of our ~1000 sampled sites was 3% and ~60% of the sites sampled had an absolute terrain slope of less than 5%. Nevertheless, some sites did contain more complex topography that could cause drastically

5   different dispersion parameters. Such sites would need to be analyzed on a case by case basis as dispersion over complex topography is usually not generalizable because every site is unique (e.g. the inverse Gaussian modeling approach might work very well at one site but poorly at another). This analysis would be non-trivial and requires high resolution topography data, surface heat flux fields and many other inputs for accurate modelling. Another possible pathway to fully investigate the effects of terrain is to investigate correlation between site emissions determined using Gaussian models and terrain slope.

10   From these analyses one can determine screening criteria to preserve data quality and examine the skill of Gaussian models over complex terrain in general. This is the subject of forthcoming work using the larger dataset and not discussed here. A final note on this topic is that the present LES represented the tanks and structures of the sites as bluff bodies that blocked the flow and created wakes, while the Gaussian models did not. This did not result in drastic difference between the Gaussian and LES results indicating that perhaps small obstructions do not have a disproportionate impact on the retrieved

15   emission rates.

5.2 Total Uncertainty Estimate

        The sources for uncertainty and bias in the Gaussian measurements discussed in this analysis are summarized in Table 56. These include the uncertainty in the Gaussian diffusion constant by comparing to LES calculated diffusion, uncertainty due to source location and height and uncertainty due to wind speed and stability class. In addition, the LES was

20   used to observe bias in the Gaussian derived concentration distributions and the controlled release was used to evaluate bias in both the Gaussian and LES results. Finally, the LES was used to determine the optimum sampling pattern to constrain actual atmospheric variability. The largest contributor to total uncertainty is atmospheric variability, the random error induced by insufficient averaging of the turbulent instantaneous plume. As atmospheric variability is impossible to separate from other sources of uncertainty, such as wind speed, it is not surprising that it is the largest source of uncertainty. As

25   described in Sect. 3.1, the LES derived atmospheric variability (defined as the standard deviation of emissions retrieved) is expected to be ~25%, considerably less than the standard deviation observed directly since the LES can capture effect of turbulence, but not due tothe effect of changes in the mean flow and meandering plumes whichthat can contribute significantly to overall atmospheric variability (Vickers, Mahrt and Belusic, 2008, Mortarini et al., 2016). This is also the practical2016). LES has a limited ability to represent the very-large scale motions (Kunkel and Marusic, 2006) or

30   some eddy features (Glendending, 1996) due to its limited horizontal domain size and idealized forcing (e.g. de Roode et al, 2004, Agee and Gluhovsky, 1999ab). In addition, the limitation of LES in the surface layer due to applications of the Monin-Obukhov similarity theory is also one of the concerns (e.g. Khanna and Brasseur, 1997). We do not intend to represent all

sources of uncertainties pertaining to the atmospheric conditions. Instead, the focus of the LES is to resolve the range of scales that are critical for the turbulent diffusion of the plume, which is often represented as a single eddy diffusivity coefficient in the Gaussian plume models. Thus, the LES predicted variability is the practical lower limit of uncertainty for this method since other sources of uncertainty could be mitigated by better on-site measurements and source location

5   detection, but some random atmospheric variability is ~~impossible~~more difficult to ~~reduce~~constrain. While the higher observed atmospheric variability may be partially explained by far more complex real world conditions leading to higher standard deviations, this also emphasizes the possibility that other sources of uncertainty are contained in this realization of atmospheric variability. For instance, most sites, while relatively flat, still have some inhomogeneous terrain that can influence and deflect wind or lead to increased turbulence. To combine the remaining sources of error, a Monte Carlo

10   simulation of errors through the Gaussian equation was performed. Inputs are available in Table 7 and an example output if Fig. S15. This approach was determined to be the most appropriate ~~as opposed to adding in quadrature~~ as uncertainties may not be normally distributed and emissions are constrained to be above zero, causing a skew to the emission retrievals. ~~The combined effects produce a skewed distribution with a standard deviation of 100%.~~A generic scenario matching average conditions experienced during the measurements was devised using a downwind distance

15   of 200m, 1.5 m s^{-1} wind speed, neutral stability and 260 ppb observation enhancement. The specifics with regards to distributions assumed for each uncertainty parameter are shown in Table 7 for the SS Gaussian and MS Gaussian approaches with 1 and 10 transects, as well as a theoretical lower limit scenario. For each scenario, 1000 randomly generated samples of $C_{Observation}$, $C_{Background}$, and C_{Model} were obtained and used according to Eq. 2 to obtain a distribution of Q samples. The Q samples were then used to estimate the 95% confidence interval. The combined effects produce a skewed distribution of

20   emission rates as shown in Fig. S15. As discussed in Sect. 5.1, this analysis focuses on relatively flat, simple sites and is not intended to be generalized to extremely complex sites. We caution that the plume diffusion uncertainty is therefore a minimum expected value and could be a major source of uncertainty for very complex sites not investigated here.

~~5.3 Uncertainty through Different Averaging Approaches~~

~~One additional consideration is the method used to average transects to derive an emission~~
25   ~~estimate. We have chosen to average the emissions deduced from individual transects. Alternatively, the multiple transects themselves could be averaged to produce a single emission estimate. The latter method should theoretically be more Gaussian in shape and more comparable to the model, but requires enough transects to produce a Gaussian profile and may not be appropriate for sites with a limited number of transects. However, we analyzed a representative site as a comparison to provide information~~
30   ~~as to what, if any, difference in retrieved emissions may be expected with this method. As an example, Site 3 was chosen to average the multiple transects previously shown in Fig. 2. The averaged plume, shown in Fig. 12, was then used to calculate the emission rate using the IGM approach, also using the model averaged across all transects. The averaged transect emission rate is 1.2 kg hr^{-1}, extremely close~~

to the 1.1 kg hr⁻¹ average of all single transects. The single transects range from passes with no plume detected to a maximum of 2.6 kg hr⁻¹.

~~to the 1.1 kg hr^{-1} average of all single transects. The single transects range from passes with no plume detected to a maximum of 2.6 kg hr^{-1}.~~

~~Averaging multiple transects has the benefit of reducing the influence of atmospheric variability on the uncertainty of the measurement; however, longer measurement time is still required due to the need for many transects. Each approach, averaging emissions vs. averaging transects, produces similar final results indicating that averaging method is not a driver of emission uncertainty. Either approach may be acceptable given the constraints and intent of a sampling session. The variability of single transects of emissions may be a useful tool for data quality control while averaged transects may be useful in additional analysis intent on pinpointing the location of an unknown source.~~

~~5.4~~5.3 Advantages and Disadvantages

Of the approaches compared in this analysis, the LES results require far more computational and processing time. Though inputs can be estimated from other sources (i.e. NOAA), we chose to measure meteorological variables directly~~,~~. which contributed to significantly longer measurement time. While this should be considered best practice when producing computationally expensive LES outputs, there are no inherent differences between inputs needed for LES and Gaussian approaches and the Gaussian approach also benefits from on-site measurements making in-field measurement time theoretically similar. In practice, the additional set-up time for meteorological instrumentation can increase total measurement time for sites intended for LES to >1 hour. Table ~~6~~8 summarizes the main advantages and disadvantages for each technique. The main advantage of the LES is the ability to directly calculate the plume diffusion rather than rely on simplified models. However, the LES simulations shown here have all been initialized for neutral conditions, which is the stability class we targeted during sampling. This is a relatively transient atmospheric phase typically only occurring in the morning and evening around sunrise and sunset and generally other stability classes are encountered. The Gaussian method enables easy corrections for different stabilities allowing quick processing of data collected in these regimes. While LES can be programmed with different stabilities, the additional computational cost is great and the ~~system~~surface energy ~~change~~budget must be known~~,~~. which introduces another source of uncertainty as actual heat fluxes may vary over the domain of interest and single point measurements may not be accurate. For these reasons, LES alone would not be the recommended method of calculating emission rates based on our study. Likewise, the single transect method, though fast, has many sources of uncertainty. Hence, the strategic combination of all of these approaches described in Sect. 3 is expected to maximize sampling efficiency while minimizing uncertainty. Less frequent higher intensity measurements (from on-site meteorological data and multiple transects) can be used to provide a better estimate of uncertainty for single transect approaches.

6 Comparison to Previous Uncertainty Estimates

Overall, we find LES to be a useful tool to examine Gaussian sampling strategy and sources of uncertainty for mobile laboratory measurements. Subsampling the LES output generates an optimum sampling pattern of at least 10 transects per site to obtain reliable statistics of measurement uncertainty due to atmospheric variability, which is the largest source of uncertainty. When sampling at distances greater than 150 m downwind of sites, the uncertainty due to source location and height are generally less than 20% (for cases where source location is known within 50 m and source height is known within 10 m). Using the LES and a controlled release, we confirm that the Gaussian model performs well ~~and~~when in-situ winds are available. The NOAA estimated winds can be a source of error, but we did not observe a systematic difference between the NOAA and in-situ winds, thus no sources of bias using our approach are ~~observable.~~expected on average. LES is therefore not required for studies where source strength calculation is the main goal and other complicating factors such as complex topography are not present. The emission retrievals generally fall within a range of two. From this we use Monte Carlo analysis to extrapolate that the 95% confidence interval for sites with standard sampling (n=2) ranges from 0.~~05x~~05q–6.~~0x~~5q where ~~x~~q is the emission rate. Using the same approach, sites that had multiple passes and wind measurements (replicate/intensive) can be further constrained to 0.~~07x –2.5x~~10q –3.0q. This uncertainty estimate is higher than Lan et al. (2015) who reported 0.~~5x~~5q–1.~~5x~~5q at the 95% confidence interval. Their study is identical to the theoretical lower limit of uncertainty we calculate by assuming only the LES predicted atmospheric variability of 25%. It should be noted that Lan et al. (2015) did incorporate some averaging over a time frame of >10 min to their measurements, which, as discussed in Section 5.3, can also decrease uncertainty. However, this would not mitigate all other sources of error previously discussed.

In addition, our standard sampling uncertainty range is greater than other Gaussian approaches with controlled releases which reported 0.~~28(x)~~28q – 3.~~6(x)~~6q and 0.~~334(x)~~334q – 3.~~34(x)~~34q (Rella et al., 2015, Yacovitch et al., 2015); the uncertainty range of the multi-transect sites studies here was similar in magnitude~~.~~ (0.05q–6.5q). However, their analysis did not account for additional sources of uncertainty (source location, stability, wind speed~~)~~), which can result in uncertainties larger than the reported values. In addition, Rella et al. (2015) incorporated vertical information to inform their Gaussian plume model creating a mass balance Gaussian hybrid, though only a small vertical profile was available (4 measurement points up to 4 m). As the vertical flux was seen to be a potential source of error in this work, measurements of this metric may reduce uncertainty. As described in Sect. 3.2, the observed atmospheric variability (observed from transect to transect emission rate variability) can range from 10-200~~% meaning that in~~%. Atmospheric variability (random error) was the single largest driver of uncertainty in the Monte Carlo simulations (see for example Fig. S15) because the model can be improved with better information and wind measurement. To reduce the uncertainty from atmospheric variability more observations are needed; however, sometimes this is impractical. In-situ observations of variability ~~and~~ from repeat

23

measurements at a single site may potentially be used as a post screening ~~out~~method to exclude conditions ~~with unacceptably~~that will lead to extremely high ~~variability may be a viable way to reduce~~ uncertainty in single transect sites. Other factors such as wind speed and stability can have a strong effect and can be quantified to reduce uncertainty.

7 Recommendations

While the uncertainty derived for mobile Gaussian techniques is large compared to many other techniques discussed in Sect. 1, it is low enough to reliably separate 'normal emissions' from 'extreme emissions' that are orders of magnitude larger. Many emission sources exhibit lognormal distributions where this condition is met, making sampling a reliable way to identify extreme emission sites. However, longer sampling time, reliable mobile wind sampling and visualization of plume distributions are needed to feasibly constrain this method to under 50% for routine measurements. In summary, to facilitate more constrained uncertainty from other mobile platform based Gaussian emission estimates, we recommend the following:

1. Sites should be isolated to reduce contamination from other sources and be accessible from thoroughfares at least 100 m away.
2. On-site wind measurement should be collected whenever possible.
3. Additional data should be collected such as photographs (used here) or IR imagery to precisely locate the sources whenever possible.
4. Ideally, all sites should use ≥ 10 sampling transects to reliably ~~measure~~constrain atmospheric variability.
5. For experiments where sampling frequency is at a premium, at least 1 site per sampling outing should be repeated with ≥ 10 sampling transects to reliably ~~measure~~constrain atmospheric variability which is expected to be the largest source of the uncertainty estimate.
6. Uncertainty analysis should be a systematic part of Gaussian sampling design.
7. In the absence of other experiments to study measurement uncertainty (controlled releases), the repeat measurements may be a suitable approximation for the *minimum* expected uncertainty.
8. While the ~~strategy~~strategies described in the study ~~was~~(see Fig. 6) were developed for well pads, the findings are generalizable to other 'point-like' sources (<2,500 m^2 and >100 m downwind) with simple terrain.

8 Data Availability

A data file containing all the emissions, LOD, uncertainty estimate, meteorology, site locations and traits including spud date, operator, production and status will be submitted to DataSpace at Princeton University (https://dataspace.princeton.edu/jspui/). This archive is free and open to the public.

9 Appendix A

Diffusion Equations (m) as presented in Zannetti (1990).

Briggs 1973 Rural

Stability	Horizontal (σ_y)	Vertical (σ_z)
A	$0.22x(1 + 0.0001x)^{-0.5}$	$0.20x$
B	$0.16x(1 + 0.0001x)^{-0.5}$	$0.12x$
C	$0.11x(1 + 0.0001x)^{-0.5}$	$0.08x(1 + 0.0002x)^{-0.5}$
D	$0.08x(1 + 0.0001x)^{-0.5}$	$0.06x(1 + 0.0015x)^{-0.5}$
E	$0.06x(1 + 0.0001x)^{-0.5}$	$0.03x(1 + 0.0003x)^{-0.5}$
F	$0.04x(1 + 0.0001x)^{-0.5}$	$0.016x(1 + 0.0003x)^{-0.5}$

Briggs 1973 Urban

Stability	Horizontal (σ_y)	Vertical (σ_z)
A	$0.32x(1 + 0.0004x)^{-0.5}$	$0.24x(1 + 0.001x)^{0.5}$
B	$0.32x(1 + 0.0004x)^{-0.5}$	$0.24x(1 + 0.001x)^{0.5}$
C	$0.22x(1 + 0.0004x)^{-0.5}$	$0.20x$
D	$0.16x(1 + 0.0004x)^{-0.5}$	$0.14x(1 + 0.0003x)^{-0.5}$
E	$0.11x(1 + 0.0004x)^{-0.5}$	$0.08x(1 + 0.00015x)^{-0.5}$
F	$0.11x(1 + 0.0004x)^{-0.5}$	$0.08x(1 + 0.00015x)^{-0.5}$

Gifford 1961

$$\frac{k_1 x}{[1 + \left(\frac{x}{k_2}\right)]^{k_3}} \qquad \frac{k_4 x}{[1 + \left(\frac{x}{k_2}\right)]^{k_5}}$$

Stability	Horizontal (σ_y)		Vertical (σ_z)		
	k_1	k_3	k_2	k_4	k_5
A	0.25	0.189	927	0.102	-1.918
B	0.202	0.162	370	0.0962	-0.101
C	0.134	0.134	283	0.0722	0.102
D	0.0787	0.135	707	0.0475	0.465
E	0.0566	0.137	1070	0.0335	0.624
F	0.0370	0.134	1170	0.022	0.7

$$\sigma = ax^b$$

Smith 1968

Stability	Horizontal (σ_y)		Vertical (σ_z)	
	a	b	a	b
A	0.4	0.91	0.41	0.91
B	0.36	0.86	0.33	0.86
C	0.36	0.86	0.33	0.86
D	0.32	0.78	0.22	0.78
E	0.32	0.78	0.22	0.78
F	0.31	0.71	0.06	0.71

10 Author Contributions

D.R.C, Q.L. and E.B.Z. implemented LES modeling for sites. D.R.C, J.M.L and H.M.L. analyzed raw data and provided Gaussian fluxes. D.R.C., J.M.L, H.M.L, J.P.F., B.B., L.M.G., X.G., J.M., D.P., L.W., and M.A.Z. collected in-situ data. D.R.C, L.M.G., J.F.P., E.B.Z. and M.A.Z. assisted in sampling design. L.M.G, D.P., B.B. and X.G. provided technical support for data collection and instrument integration. J.M. provided mechanical assistance in mobile lab design and deployment. D.R.C prepared the manuscript with feedback from all authors.

~~10~~11 Competing Interests

The authors declare that they have no conflict of interest.

1213 References

Agee, E., and Gluhovsky, A.: LES Model Sensitivities to Domains, Grids, and Large-Eddy Timescales, J. Atmos. Sci., 56, 599-605, doi: 10.1175/1520-0469(1999)056<0599:LMSTDG>2.0.CO;2, 1999a.

Agee, E., and Gluhovsky, A.: Further Aspects of Large Eddy Simulation Model Statistics and Inconsistencies with Field Data, J. Atmos. Sci., 56, 2948-2950, doi: 10.1175/1520-0469(1999)056<2948:FAOLES>2.0.CO;2, 1999b.

Albertson, J. D., Harvey, T., Foderaro, G., Zhu, P., Zhou, X., Ferrari, S., Amin, M. S., Modrak, M., Brantley, H., and Thoma, E. D.: A Mobile Sensing Approach for Regional Surveillance of Fugitive Methane Emissions in Oil and Gas Production, Envir. Sci. Tech., 50, 2487-2497, doi: 10.1021/acs.est.5b05059, 2016.

Allen, D. T., Torres, V. M., Thomas, J., Sullivan, D. W., Harrison, M., Hendler, A., Herndon, S. C., Kolb, C. E., Fraser M. P., Hill, A. D., and Lamb, B. K.: Measurements of methane emissions at natural gas production sites in the United States, P. Natl. Acad. Sci. U.S.A., 110, 17768-73, doi: 10.1073/pnas.1304880110, 2013.

Allen, D. T., Pacsi, A. P., Sullivan, D. W., Zavala-Araiza, D., Harrison, M., Keen, K., Fraser, M. P., Hill, A. D., Sawyer, R. F., and Seinfeld, J. H.: Methane emissions from process equipment at natural gas production sites in the United States: Pneumatic controllers, Envir. Sci. Tech., 49, 633-40, doi: 10.1021/es5040156, 2014.

Baker, L. H., Collings, W. J., Olivie, D. J. L., Cherian, R., Hodnebrog, O., Myhre, G., and Quass, J.: Climate responses to anthropogenic emissions of short-lived climate pollutants, Atmos. Chem. Phys., 15, 8201-8216, doi: 10.5194/acp-15-8201-2015, 2015.

Batchelor, G. K.: Diffusion in a field of homogeneous turbulence. I. Eulerian Analysis, Aust. J. Sci. Res., 2, 437-50, doi: 10.1071/CH9490437, 1949.

Black, T. L.: The New NMC Mesoscale Eta Model: Description and Forecast Examples, Weather Forecast., 9, 265-278, doi: 10.1175/1520-0434(1994)009<0265:TNNMEM>2.0.CO;2, 1994.

Bosanquet, C. H., and Pearson, J. L.: The Spread of Smoke and Gases from Chimneys, Trans. Faraday Soc., 32, 1249-63, 1936.

Bou-Zeid, E., Meneveau, C., and Parlange, M.: A Scale-Dependent Lagrangian Dynamic Model for Large Eddy Simulation of Complex Turbulent Flows, Phys. Fluids, 17, doi:10.1063/1.1839152, 2005.

Bowerman, N. H. A., Frame, D. J., Huntingford, C., Lowe, J. A., Smith, S. M., and Allen, M. R.: The role of short-lived climate pollutants in meeting temperature goals, Nat. Clim. Change, 3, 1021-24, doi: 10.1038/nclimate2034, 2013.

Brandt, A. R., Heath, G. A., and Cooley, D.: Methane Leaks from Natural Gas Systems Follow Extreme Distributions, Envir. Sci. Tech., 50, 12512-12520, doi: 10.1021/acs.est.6b04303, 2016.

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, Envir. Sci. Tech., 48, 14508-15, doi: 10.1021/es503070q, 2014.

Briggs, G. A.: Diffusion estimation for small emissions, in environmental research laboratories, Air Resources Atmospheric Turbulence and Diffusion Laboratory 1973 Annual Report, USAEC Report ATDL-106, National Oceanic and Atmospheric Administration, 83-145,1973.

Caulton, D. R., Shepson, P. B., Santoro, R. L., Sparks, J. P., Howarth, R. W., Ingraffea, A. R., Cambaliza, M. O., Sweeney, C., Karion, A., Davis, K. J., and Stirm, B. H.: Toward a better understanding and quantification of methane emissions from shale gas development, P. Natl. Acad. Sci. U.S.A., 111, 6237-42, doi: 10.1073/pnas.1316546111 , 2014.

Center for Chemical Profess Safety: Averaging Times, Concentration Fluctuations, and Modeling Uncertainties, in: Guidelines for Use of Vapor Cloud Dispersion Models, 2nd ed., John Wiley & Sons, Inc., New York, U.S.A., 1996.

Deardorff, J. W. and Willis, G. E.: Concentration Fluctuations Within a Laboratory Convectively Mixed Layer, American Meteorological Society, Boston, MA, U.S.A., 357-383, 1988.

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, Atmos. Environ., 115, 101-9, doi: 10.1016/j.atmosenv.2015.05.042, 2015.

Frankenberg, C., Thorpe, A. K., Thompson, D. R., Hulley, G., Kort, E. A., Vance, N., Borchardt, J., Krings, T., Gerilowski, K., Sweeney, C., and Conley, S.: Airborne methane remote measurements reveal heavy-tail flux distribution in Four Corners region, P. Natl. Acad. Sci. U.S.A., 113, 9734-39, doi: 10.1073/pnas.1605617113, 2016.

Fritz, B. K., Shaw, B. W., Parnell, C. B.: Influence of Meteorological Time Frame and Variation on Horizontal Dispersion Coefficients in Gaussian Dispersion Modeling, Trans. A.S.A.E., 48, 1185-1196, doi: 10.13031/2013.18501, 2005.

Gålfalk, M., Olofsson, G., Crill, P., and Bastviken, D.: Making methane visible, Nat. Clim. Change, 6, 426-30, doi:10.1038/nclimate2877, 2016.

Gifford, Jr., F. A.: An outline of theories of diffusion in the lower layers of the atmosphere, in: Meteorology and Atomic Energy, Slade, D. (Ed.), U.S. Atomic Energy Commission, TID-24190, Springfield, Virginia, 65-116, 1968.

Gifford, Jr., F. A.: Turbulent diffusion-typing schemes: a review, Nucl. Safety, 17, 68-86, 1976.

Glendening, J. W.: Lineal Eddy Features under Strong Shear Conditions, J. Atmos. Sci., 53, 3430-3449, doi: 10.1175/1520-0469(1996)053<3430:LEFUSS>2.0.CO;2, 1996.

Goetz, J. D., Floerchinger, C., Fortner, E. C., Wormhoundt, J., Massoli, P., Knighton, W. B., Herndon, S. C., Kolb, C. E., Knipping, E., Shaw, S. L., and DeCarlo, P. F.: Atmospheric Emission Characterization of Marcellus Shale Natural Gas Development Sites, Envir. Sci. Tech., 49, 7012-7020, doi: 10.1021/acs.est.5b00452, 2015.

Hilst, G. R.: The Dispersion of Stack Gases in Stable Atmospheres, J. Air Pollut. Control Assoc., 7, 205-10, doi: 10.1080/00966665.1957.10467804, 1957.

Kang, M., Kanno, C. M., Reid, M. C., Zhang, X., Mauzerall, D. L., Celia, M. A., Chen, Y., and Onstott, T. C.: Direct measurements of methane emissions from abandoned oil and gas wells in Pennsylvania, P. Natl. Acad. Sci. U.S.A., 111, 18173-7, doi: 10.1073/pnas.1408315111, 2014.

Karion, A., Sweeney, C., Pétron, G., Frost, G., Hardesty, R. M., Kofler, J., Miller, B. R., Newberger, T., Wolter, S., Banta, R., and Brewer, A.: Methane emissions estimate from airborne measurements over a western United States natural gas field, Geophys. Res. Lett. 40, 4393-7, doi: 10.1002/grl.50811, 2013.

Karion, A., Sweeney, C., Kort, E. A., Shepson, P. B., Brewer, A., Cambaliza, M., Conley, S. A., Davis, K., Deng, A., Hardesty, M., and Herndon, S. C.: Aircraft-based estimate of total methane emissions from the Barnett Shale region. Envir. Sci. Tech., 49, 8124-31, doi: 10.1021/acs.est.5b00217, 2015.

Khanna, S., and Brasseur, J. G.: Analysis of Monin-Obukhov similarity from large-eddy simulation, J. Fluid Mech., 345, 251-286, doi:10.1017/S0022112097006277 , 1997.

Kirschke, S., Bousquet, P., Ciais, P., Saunois, M., Canadell, J. G., Dlugokencky, E. J., Bergamaschi, P., Bergmann, D., Blake, D. R., Bruhwiler, L., Cameron-Smith, P., Castaldi, S., Chevallier, F., Feng, L., Fraser, A., Heimann, M., Hodson, E. L., Houweling, S., Josse, B., Fraser, P. J., Krummel, P. B., Lamarque, J.-F., Langenfelds, R. L., Le Quéré, C., Naik, V., O'Doherty, S., Palmer, P. I., Pison, I., Plummer, D., Poulter, B., Prinn, R. G., Rigby, M., Ringeval, B., Santini, M., Schmidt, M., Shindell, D. T., Simpson, I. J., Spahni, R., Steele, L. P., Strode, S. a., Sudo, K., Szopa, S., van der Werf, G. R., Voulgarakis, A., van Weele, M., Weiss, R. F., Williams, J. E. and Zeng, G.: Three decades of global methane sources and sinks, Nat. Geosci., 6(10), 813–823, doi:10.1038/ngeo1955, 2013.

Kort, E. A., Frankenberg, C., Costigan, K. R., Lindenmaier, R., Dubey, M. K., and Wunch, D.: Four corners: The largest US methane anomaly viewed from space, Geophys. Res. Lett., 41, 6898-903, doi: 10.1002/2014GL061503, 2014.

Kuai, L., Worden, J. R., Li, K.-F., Hulley, G. C., Hopkins, F. M., Miller, C. E., Hook, S. J., Duren, R. M., and Aubrey, A. D.: Characterization of anthropogenic methane plumes with the Hyperspectral Thermal Emission Spectrometer (HyTES): a retrieval method and error analysis. Atmos. Meas. Tech., 9, 3165, 2016.

Kunkel, G. J., and Marusic, I.: Study of the near-wall-turbulent region of the high-Reynolds-number boundary layer using an atmospheric flow. J. Fluid Mech., 548, 375-402, doi: 10.1017/S0022112005007780, 2006.

Lamb, B. K., Edburg, S. L., Ferrara, T. W., Howard, T., Harrison, M. R., Kolb, C. E., Townsend-Small, A., Dyck, W., Possolo, A., and Whetstone, J. R.: Direct measurements show decreasing methane emissions from natural gas local distribution systems in the United States, Envir. Sci. Tech., 49, 5161-9, doi: 10.1021/es505116p, 2015.

Lan, X., Talbot, R., Laine, P., and Torres, A.: Characterizing fugitive methane emissions in the Barnett Shale area using a mobile laboratory, Envir. Sci. Tech., 49, 8139-46, doi: 10.1021/es5063055, 2015.

Lavoie, T. N., Shepson, P. B., Cambaliza, M. O., 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, Envir. Sci. Tech., 49, 7904-13, doi: 10.1021/acs.est.5b00410, 2015.

Li, Q., Bou-Zeid, E., and Anderson, W.: The Impact and Treatment of the Gibbs Phenomenon in Immersed Boundary Method Simulations of Momentum and Scalar Transport, J. Comput. Phys., 310, 237–51, doi:10.1016/j.jcp.2016.01.013, 2016.

Li, Q., Bou-Zeid, E., Anderson, W., Grimmond S., and Hultmark M.: Quality and Reliability of LES of Convective Scalar
    Transfer at High Reynolds Numbers, Int. J. Heat Mass Transfer, 102, 959–970,
    doi:10.1016/j.ijheatmasstransfer.2016.06.093, 2016.

Mason, P. J.: Large-Eddy Simulation of Dispersion in Convective Boundary Layers with Wind Shear, Atmos. Environ., 26A,
5   1561-1571, doi: 10.1016/0960-1686(92)90056-Q, 1992.

McDermitt, D., Burba, G., Xu, L., Anderson, T., Komissarov, A., Riensche, B., Schedlbauer, J., Starr, G., Zona, D., Oechel,
    W., Oberbayer, S., and Hastings, S.: A new low-power, open-path instrument for measuring methane flux by eddy
    covariance, Appl. Phys. B, 102, 391-405, doi: 10.1007/s00340-010-4307-0, 2010.

Metzger, M., McKeon, B. J., and Holmes, H.: The near-neutral atmospheric surface layer: turbulence and non-stationarity,
10   Phil. Trans. R. Soc. A, 365, 859-876, doi: 10.1098/rsta.2006.1946, 2007.

Mortarini, L., Stefanello, M., Degrazia, G., Roberti, D., Castelli, S. T., and Anfossi, D.: Characterization of Wind
    Meandering in Low-Wind-Speed Conditions, Bound.-Lay. Meteorol., 161, 165-82, doi: 10.1007/s10546-016-
    0165-6, 2016.

Mittal, R., and Iaccarino, G.: Immersed Boundary Methods, Annu. Rev. Fluid Mech., 37, 239–61,
15   doi:10.1146/annurev.fluid.37.061903.175743, 2005.

Nathan, B. J., Golston, L. M., O'Brien, A. S., Ross, K., Harrison, W. A., Tao, L., Lary, D. J., Johnson, D. R., Covington, A.
    N., Clark, N. N., and Zondlo, M. A.: Near-field characterization of methane emission variability from a compressor
    station using a model aircraft, Envir. Sci. Tech., 49, 7896-903, doi: 10.1021/acs.est.5b00705, 2015.

Nickels, T. B., Marusic, I., Hafez, S., Hutchins, N. and Chong, M. S.: Some predictions of the attached eddy model for a
20   high Reynold number boundary layer, Phil. Trans. R. Soc. A, 365, 807-822 , doi: 10.1098/rsta.2006.1950, 2007.

Nieuwstadt, F. T. M., and de Valk, J. P.J. M. M.: A Large Eddy Simulation of Buoyant and Non-Buoyant Plume Dispersion
    in the Atmospheric Boundary Layer, Atmos. Environ., 21, 2573-2587, doi: 10.1016/0004-6981(87)90189-2,1987.

Omara, M., Sullivan, M. R., Li, X., Subramanian, R., Robinson, A. L., and Presto, A. A.: Methane emissions from
    conventional and unconventional natural gas production sites in the Marcellus Shale Basin, Envir. Sci. Tech., 50,
25   2099-107, doi: 10.1021/acs.est.5b05503, 2016.

Peischl, J., Ryerson, T. B., Brioude, J., Aikin, K. C., Andrews, A. E., Atlas, E., Blake, D., Daube, B. C., Gouw, J. A.,
    Dlugokencky, E., and Frost, G. J.: Quantifying sources of methane using light alkanes in the Los Angeles basin,
    California. J. Geophys. Res.-Atmos., 118, 4974-90, doi: 10.1002/jgrd.50413, 2013.

Peischl, J., Ryerson, T. B., Aikin, K. C., Gouw, J. A., Gilman, J. B., Holloway, J. S., Lerner, B. M., Nadkarni, R., Neuman,
30   J. A., Nowak, J. B., and Trainer, M. Quantifying atmospheric methane emissions from the Haynesville, Fayetteville,
    and northeastern Marcellus shale gas production regions, J. Geophys. Res.-Atmos., 120, 2119-39, doi:
    10.1002/2014JD022697, 2015.

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., and Trainer, M.: Quantifying atmospheric methane emissions from oil and natural gas production in

the Bakken shale region of North Dakota. J. Geophys. Res.-Atmos., 121, 6101-11, doi: 10.1002/2015JD024631, 2016.

Perry, A. E., and Li., J. D.: Experimental support for the attached-eddy hypothesis in zero-pressure-gradient turbulent boundary layers, J. Fluid Mech., 218, 405-438, doi: 10.1017/S0022112090001057, 1990.

Pétron, G., Frost, G., Miller, B. R., Hirsch, A. I., Montzka, S. A., Karion, A., Trainer, M., Sweeney, C., Andrews, A. E., Miller, L., and Kofler, J.: Hydrocarbon emissions characterization in the Colorado Front Range: A pilot study, J. Geophys. Res.-Atmos., 117, D04304, doi: 10.1029/2011JD016360, 2012.

Pétron, G., Karion, A., Sweeney, C., Miller, B. R., Montzka, S. A., Frost, G. J., Trainer, M., Tans, P., Andrews, A., Kofler, J., and Helmig, D.: A new look at methane and nonmethane hydrocarbon emissions from oil and natural gas operations in the Colorado Denver-Julesburg Basin. J. Geophys. Res.-Atmos., 119, 6836-52, doi: 10.1002/2013JD021272, 2014.

Ravikumar, A. P., Wang, J., and Brandt, A. R.: Are Optical Gas Imaging Technologies Effective for Methane Leak Detection?,   Envir. Sci. Technol, 51, 718-24, doi: 10.1021/acs.est.6b03906, 2017.
Ren, X., Hall, D. L., Vinciguerra, T., Benish, S. E., Stratton, P. R., Ahn, D., Hansford, J. R., Cohen, M. D., Sahu, S., He, H.,   and Grimes, C.: Methane Emissions from the Marcellus Shale in Southwestern Pennsylvania and Northern West   Virginia Based on Airborne Measurements, J. Geophys. Res.-Atmos., 122, 4639-53, doi: 10.1002/2016JD026070,   2017.

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, Envir. Sci. Tech., 49, 4742-8, doi: 10.1021/acs.est.5b00099, 2015.

Rolph, G., Stein, A., and Stunder, B.: Real-time Environmental Applications and Display sYstem: READY, Environ. Modell. Softw., 95, 210-228, doi: 10.1016/j.envsoft.2017.06.025, 2017.

de Roode, S. R., Duynkerke, P. G., and Jonker, H. J. J.: Large-Eddy Simulation: How Large is Large Enough?, J. Atmos. Sci., 61, 403-421, doi: 10.1175/1520-0469(2004)061<0403:LSHLIL>2.0.CO;2, 2004.

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., and Robinson, A. L.: Measurements of methane emissions from natural gas gathering facilities and processing plants: measurement methods, Atmos. Meas. Tech., 8, 2017-35, doi: 10.5194/amt-——8-2017-2015, 2015.

Salesky, S. T., and Chemecki, M.: Random Errors in Turbulence Measurements in the Atmospheric surface Layer: Implications for Monin-Obukhov Similarity Theory, J. Atmos. Sci., 69, 3700-3714, doi: 10.1175/JAS-D-12-096.1, 2012.

Seinfeld, J. H. and Pandis, S. N.: Atmospheric Chemistry and Physics. John Wiley & Sons, Inc. Hoboken, New Jersey, U.S.A., 828-896, 1998.

Shah, S. and Bou-Zeid, E.: Very-Large-Scale Motions in the Atmospheric Boundary Layer Educed by Snapshot Proper Orthogonal Decomposition, Bound.-Lay. Meteorol., 153, 355-387, doi: 10.1007/s10546-014-9950-2, 2014.

Smith, M. E.: Recommended guide for the prediction of the dispersion of airborne effluents. ASME, New York, 1968.

Subramanian, R., Williams, L. L., Vaughn, T. L., Zimmerle, D., Roscioli, J. R., Herndon, S. C., Yacovitch, T. I., Floerchinger, C., Tkacik, D. S., Mitchell, A. L., and Sullivan, M. R.: Methane Emissions from Natural Gas Compressor Stations in the Transmission and Storage Sector: Measurements and Comparisons with the EPA Greenhouse Gas Reporting Program Protocol, Envir. Sci. Tech., 49, 3252-61, doi: 10.5194/amt-8-2017-2015, 2015.

Sutton, O. G.: A Theory of Eddy Diffusion in the Atmosphere, P. R. Soc. Lond. A-Conta., 135, 143-165, 1932.

Tao, L., Sun, K., Miller, D. J., Pan, D., Golston, L. M., and Zondlo, M. A.: Low-power, open-path mobile sensing platform for high-resolution measurements of greenhouse gases and air pollutants, Appl. Phys. B-Lasers O., 119, 153-64, doi: 10.1007/s00340-015-6069-1, 2015.

Thorpe, A. K., Frankenberg, C., Aubrey, A. D., Roberts, D. A., Nottrott, A. A., Rahn, T. A., Sauer, J. A., Dubey, M. K., Costigan, K. R., Arata, C., and Steffke, A. M.: Mapping methane concentrations from a controlled release experiment using the next generation Airborne Visible/Infrared Imaging Spectrometer (AVIRIS-NG), Remote Sens. Environ., 179, 104-15, doi: 10.1016/j.rse.2016.03.032, 2016.

Townsend, A. A.: Equilibrium layers and wall turbulence, J. Fluid. Mech. 11, 97-120, doi:10.1017/S0022112061000883, 1961.

Townsend-Small, A., Marrero, J. E., Lyon, D. R., Simpson, I. J., Meinardi, S., and Blake, D. R.: Integrating source apportionment tracers into a bottom-up inventory of methane emissions in the Barnett Shale hydraulic fracturing region, Envir. Sci. Tech., 49, 8175-82, doi: 10.1021/acs.est.5b00057, 2015.

Tseng, Y. H., Meneveau, C., and Parlange, M. B.: Modeling Flow Around Bluff Bodies and Predicting Urban Dispersion Using Large Eddy Simulation, Envir. Sci. Tech., 40, 2653–62, doi:10.1021/es051708m, 2006.

U.S. EPA: Other Test Methods (OTM) 33 and 33A Geospatial Measurement of Air Pollution-Remote Emissions Quantification Direct Assessment (GMAP-REQ-DA), https://www3.epa.gov/ttnemc01/prelim/otm33a.pdf, 2014.

Veigele, W. J., and Head, J. H.: Derivation of the Gaussian Plume Model, J. Air Pollut. Control Assoc., 28, 1139-40, doi: 10.1080/00022470.1978.10470720, 1978.

Venkatram, A. and Wyngaard, J. C (eds.).: Lectures on Air Pollution Modeling, American Meteorological Society, Boston, Massachusetts, U.S.A., 1988.

Vickers, D., Mahrt, L., and Belusic, D.: Particle simulations of dispersion using observed meandering and turbulence, Acta Geophys., 56, 234-56, doi:10.2478/s11600-007-0041-3 , 2008.

Weil, J. C.: A Diagnosis of the Asymmetry in Top-Down and Bottom-Up Diffusion Using a Lagrangian Stochastic Model, J. Atmos. Sci., 47, 501-515, doi: 10.1175/1520-0469(1990)047<0501:ADOTAI>2.0.CO;2, 1990.

Weil, J. C., Sullivan, P. P., and Moeng, C.-H.: The Use of Large-Eddy Simulations in Lagrangian Particle Dispersion Models, J. Atmos. Sci., 61, 2877-2887, doi: 10.1175/JAS-3302.1, 2004.

Woodcock, J. D., and Marusic, I.: The statistical behaviour of attached eddies, Phys. Fluids, 27, 015104, doi: 10.1063/1.4905301, 2015.

Wyngaard, J. C., and Weil, J. C.: Transport asymmetry in skewed turbulence, Phys. Fluids, 3, 155-162, doi: 10.1063/1.857874, 1991.

Yacovitch, T. I., Herndon, S. C., Petron, G., Kofler, J., Lyon, D., Zahniser, M. S., and Kolb, C. E.: Mobile Laboratory Observations of Methane Emissions in the Barnett Shale Region, Envir. Sci. Tech., 49, 7889-95, doi:10.1021/es506352j, 2015.

Zannetti, P.: Air Pollution Modeling: Theories, Computational Methods, and Available Software, Van Nostrand Reinhold, New York, U.S.A., 141-183,1990.

Zimmerle, D. J., Williams, L. L., Vaughn, T. L., Quinn, C., Subramanian, R., Duggan, G. P., Willson, B., Opsomer, J. D., Marchese, A. J., Martinez, D. M., and Robinson, A. L.: Methane emissions from the natural gas transmission and storage system in the United States, Envir. Sci. Tech., 49, 9374-83, doi: 10.1021/acs.est.5b01669, 2015.

Zickfeld, K., Solomon, S., and Gilford, D. M.: Centuries of thermal sea-level rise due to anthropogenic emissions of short-lived greenhouse gases, P. Natl. Acad. Sci. U.S.A., 114, 657-62, doi: 10.1073/pnas.1612066114, 2017.

Tables

Table 1. Comparison of CH$_4$ emission measurement techniques and reported uncertainties.

Technique	Referenced Study	Reference Uncertainty Range*
Ground-Based Thermal Imaging	Gålfalk et al., 2016	3-15%
Aircraft Remote Sensing	Kuai et al., 2016, Frankenberg et al., 2016, Thorpe et al., 2016	5-20%
Satellite Remote Sensing	Kort et al., 2014	15%
Chamber Sampling	Allen et al., 2013 and 2014, Kang et al., 2014	20-30%
Ground-Based Tracer Correlation	Lamb et al., 2015, Roscioli et al., 2015, ~~Subrumanian~~Subramanian et al., 2015, Zimmerle et al., 2015, Omara et al., 2016	20-50%
Aircraft/UAV Mass Balance	Karion et al., 2013 and 2015, Peischl et al., 2013, 2015 and 2016, Caulton et al., 2014, Pétron et al., 2014, Lavoie et al., 2015, Nathan et al., 2015~~, Ren et al., 2017~~	20-75%
Ground-Based Stationary Dispersion	Brantley et al., 2014, Foster-Wittig et al., 2015	25-60%
Tall Tower Monitoring	Pétron et al., 2012	50-100%
Ground-Based Mobile Dispersion	Lan et al., 2015, Rella et al., 2015, Yacovitch et al., 2015	50-350%

* Uncertainty range reflects author reported uncertainty on *emission* numbers, not necessarily measurement uncertainty.

Some ~~author's~~authors specify a 95% confidence interval, others use 1 or 2 standard deviations and others compute upper

5   and lower bounds.

Table 2. Summary of acquired data and data level

Sample Type	No. of Samples	Average No. of Transects	Measurement Time (min)	Wind Source	Emission Calculation Technique	Level
Standard Sample	940	2	1-5	NOAA	IGM	1
Replicate Sample	53	10	5-15	NOAA, On-site	IGM	2
Intensive Sample	17	20	15-60+	On-site	IGM, LES	3

Table 3. Summary of LES site conditions and parameters.

Characteristic	Site 1	Site 2	Site 3	Site 4	Site 5*
Sample Date	7/15/2015	7/21/2015	6/21/2016	6/23/2016	8/5/2016
Time (EDT)	18:15-20:15	7:00-9:15	9:45-10:30	20:00-20:30	16:15-20:00
No. of Transects	26	28	21	22	81
Wind Speed (m s^{-1})	1.30 ± 0.64	1.30 ± 0.70	1.9 ± 2.2	0.62 ± 0.27	1.70 ± 0.89
Wind Direction	325 ± 101	252 ± 40	276 ± 50	40 ± 47	218 ± 28
Friction Velocity (m s^{-1})	0.26	0.29	0.48	0.18	0.10-0.21
Distance from Source (m)	176	154	29	49	110
Elevation gain (m)	3.5	1.5	2	2	2
Simulation Stability	D	D	D	D	D
Simulation Time (min)	59.2	33.33	33.33	33.33	30
Simulation Warm-up Time (min)	10	5	5	5	5
Simulation x-dimension (m) (x resolution)	288 (2)	256 (1)	256 (1)	256 (1)	256 (1)
Simulation y-dimension (m) (y resolution)	256 (2)	256 (1)	256 (1)	256 (1)	256 (1)
Simulation z-dimension (m) (z resolution)	100 (1)	100 (1)	100 (1)	100 (1)	33.33 (0.2222)

* Controlled release site.

Table 4. Summary of mean emissions from three scenarios for the controlled release (± ~~1σ~~1 std. dev. of mean).

Scenario	Release 1 (kg hr⁻¹)	Release 2 (kg hr⁻¹)	Release 3 (kg hr⁻¹)
Release Rate	0.97 ± 0.01	0.~~22~~216 ± 0.002	0.~~09~~090 ± 0.002
No. of Transects	19	10	13
Gaussian w/ NOAA winds	0.97 ± 0.~~76~~17	0.79 ± 0.~~40~~13	0.35 ± 0.~~15~~04
Gaussian w/ tower winds	0.72 ± 0.~~54~~12	0.23 ± 0.~~12~~04	0.10 ± 0.~~04~~01
~~LES~~Large Eddy Simulation	0.94 ± 0.~~76~~17	0.44 ± 0.~~20~~06	0.22 ± 0.~~11~~03

Table 35. Comparison of mean emission rates using three ~~scenario~~scenarios from four sites (± ~~1σ~~1 std. dev. of mean).

Scenario	Site 1 (kg hr^{-1})	Site 2 (kg hr^{-1})	Site 3 (kg hr^{-1})	Site 4 (kg hr^{-1})
No. of Transects	26	28	21	22
~~SS~~Single Source Gaussian	1.0 ± ~~1.4~~0.3	0.19 ± 0.~~14~~03	1.~~08~~09 ± 0.~~76~~16	0.19 ± 0.~~26~~06
~~MS~~Multi-Source Gaussian	2.2 ± ~~4.7~~0.9	0.24 ± 0.~~31~~06	1.8 ± ~~1.2~~0.3	0.29 ± 0.~~27~~06
~~LES~~Large Eddy Simulation	1.5 ± ~~2.2~~0.4	0.18 ± 0.~~15~~03	0.76 ± 0.~~50~~11	0.18 ± 0.~~16~~03

Table 4. Comparison of emissions for different source location scenarios.

Site	Scenario	Emission (kg hr^{-1})	Change in x distance (m)	Change in y distance (m)	Change in z distance (m)	Difference (%)
1	1	0.54	—	—	—	—
	2	1.33	40	0	1	150
	3	0.47	0	15	7	13
2	1	0.14	—	—	—	—
	2	0.15	58	49	1	7.5
	3	0.14	72	0	7	5

~~Table 5.~~

41

Table 6. Sources and magnitude of uncertainty.

Uncertainty Source	Notes	Expected Uncertainty
Atmospheric Variability	Requires ≥10 transect to quantify, not independent from other uncertainty sources	77%
Instrumental Uncertainty	LI 7700 ~~stated~~expected 1Hz precision of ~~5~~1.6 ppb	≤1%
Plume Turbulent Diffusion	Potentially source of uncertainty and bias	25%
Source Height	Z uncertainty low at distances >150 m	15%
Source Location	Combined x~~, y~~ and ~~z~~y uncertainty low at distance >150m	~~15~~20%
Stability	Potential 1 stability class discrepancy	40%
Wind Speed	Uncertainty scales linearly	50%
~~Total~~Background	Important for small concentration enhancements	~~100~~5%

42

Table 67. Monte Carlo inputs for each uncertainty estimate.

Uncertainty (Assumed Distribution)	SS Gaussian	MS Gaussian	SS Gaussian	MS Gaussian	Lower Limit
No. of Transects	1	1	10	10	1
Source location in x-direction (Gaussian)	1 sigma = 16.7 m	1 sigma = 0	1 sigma = 16.7 m	1 sigma = 0	1 sigma = 0
Source height (uniform)	1-8 m	1 m	1-8 m	1 m	1 m
Atmospheric variability in observation (Gaussian)	1 sigma = 100% (260 ppb)	1 sigma = 100% (260 ppb)	1 sigma = 100% (260 ppb)	1 sigma = 100% (260 ppb)	1 sigma = 25% (65 ppb)
Wind speed (Gaussian)	1 sigma = 50% (0.75 m s^{-1})	1 sigma = 0	1 sigma = 50% (0.75 m s^{-1})	1 sigma = 0	1 sigma = 0
Stability (uniform)	C-E	D	C-E	D	D
Background (Gaussian)	1 sigma = 5 ppb	1 sigma = 5 ppb	1 sigma = 5 ppb	1 sigma = 5 ppb	1 sigma = 0 ppb
Uncertainty Range (95% CI)	**0.05q-6.5q**	**0.08q-3.2q**	**0.5q-2.7q**	**0.6q-1.6q**	**0.5q-1.5q**

Table 8. Summary of advantages and disadvantages of each technique.

	Measurement Time	Processing Time	Sources of Uncertainty
Single Transect Gaussian	Short (<10 min)	Short (~few minutes)	Many sources, atmospheric variability unconstrained
Multi-Transect Gaussian	Moderate (15-30 min)	Short (~few minutes)	Atmospheric variability constrained, diffusion unconstrained
Multi-Transect LES	Moderate to Long (15 min-1 hr)	Very Long (several days)	Atmospheric variability constrained, diffusion constrained

Figures
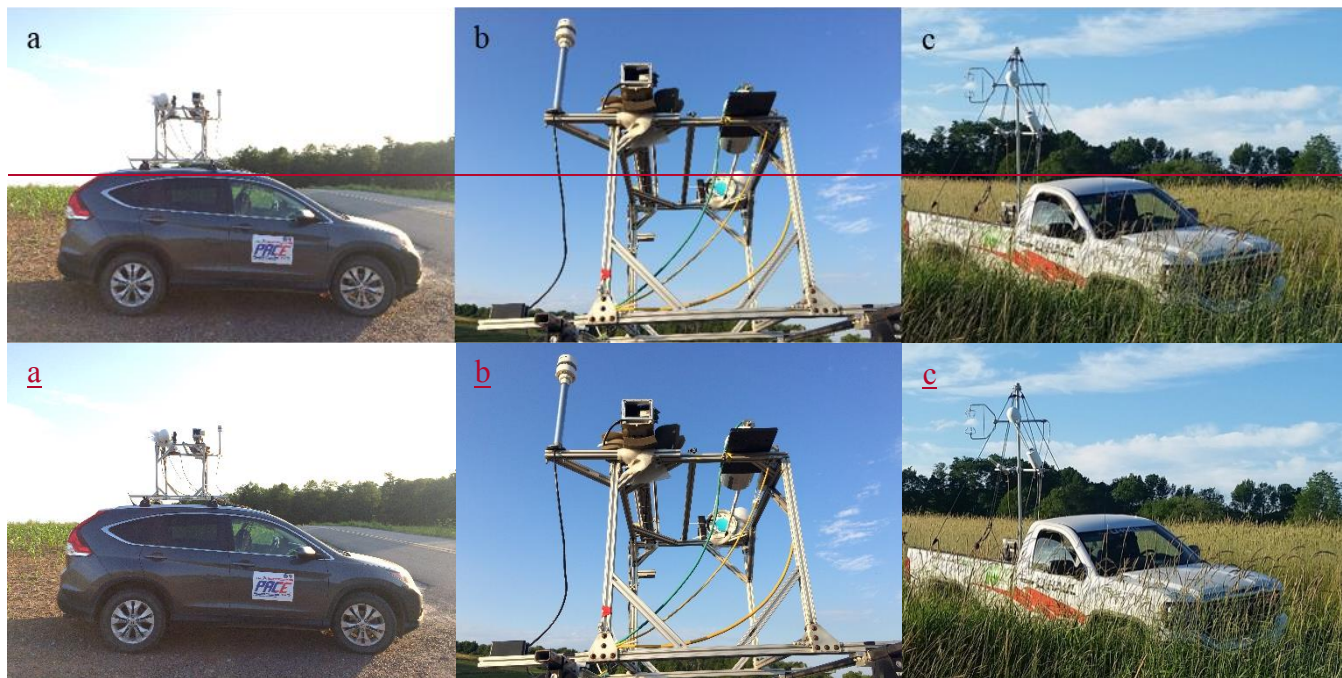
5   **Figure 1. (a) The Princeton atmospheric chemistry experiment (PACE), (b) close-up of PACE roof rack with instrumentation and (c) mobile tower platform.**
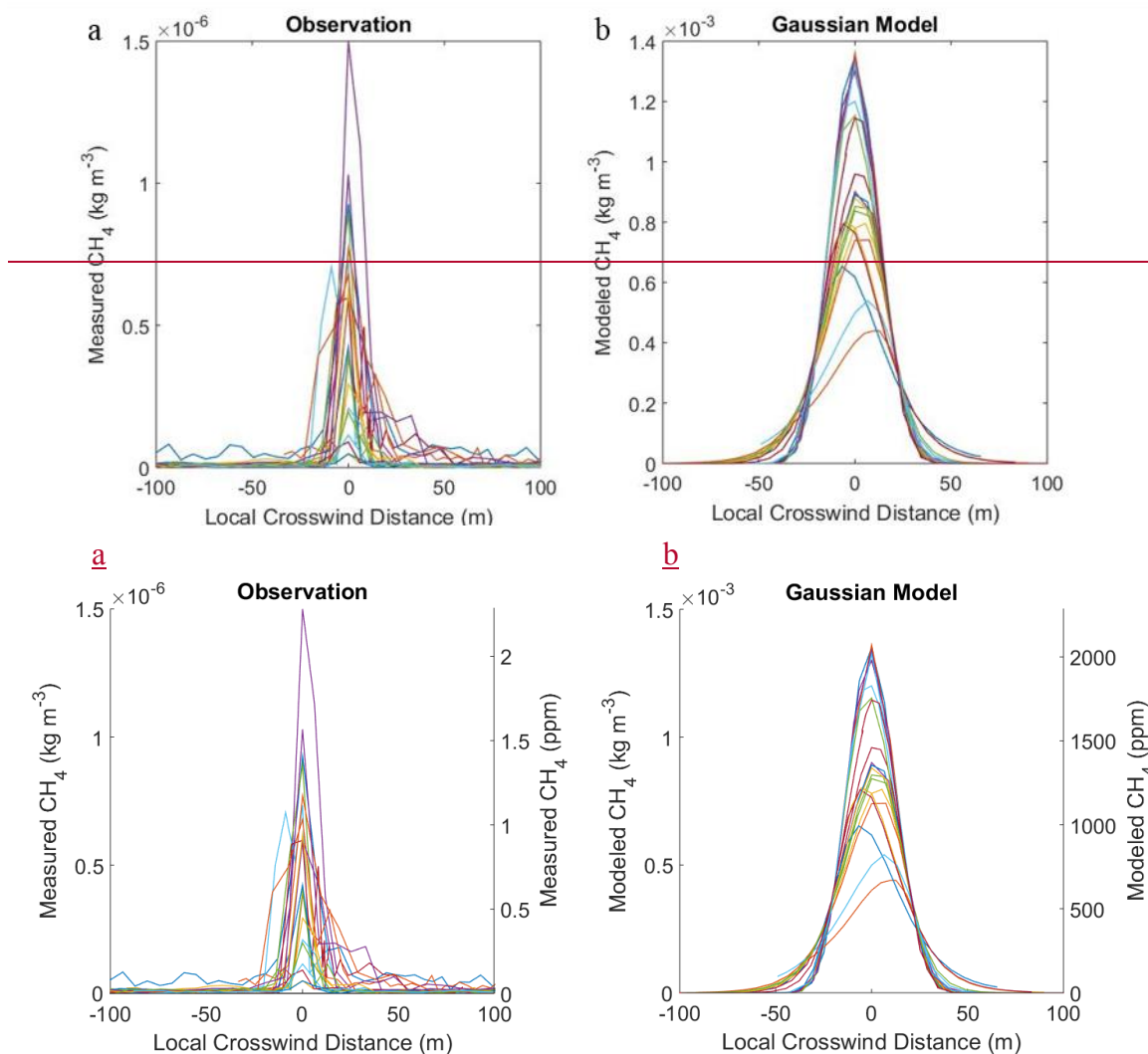
Figure 2. (a) Observations of CH₄ at a site showing multiple downwind transects ~~and~~versus the local crosswind (y) distance. (b) Gaussian outputs along the same downwind transects. Note that the scale difference between the panels is due to the inverse methodology used where the model is set to a reference emission rate (of 1 kg s⁻¹).

Site 1 Layout   Site 2 Layout   Site 3 Layout

Site 4 Layout   Site 5 Layout

Site 1 Layout   Site 2 Layout   Site 3 Layout

Site 4 Layout   Site 5 Layout
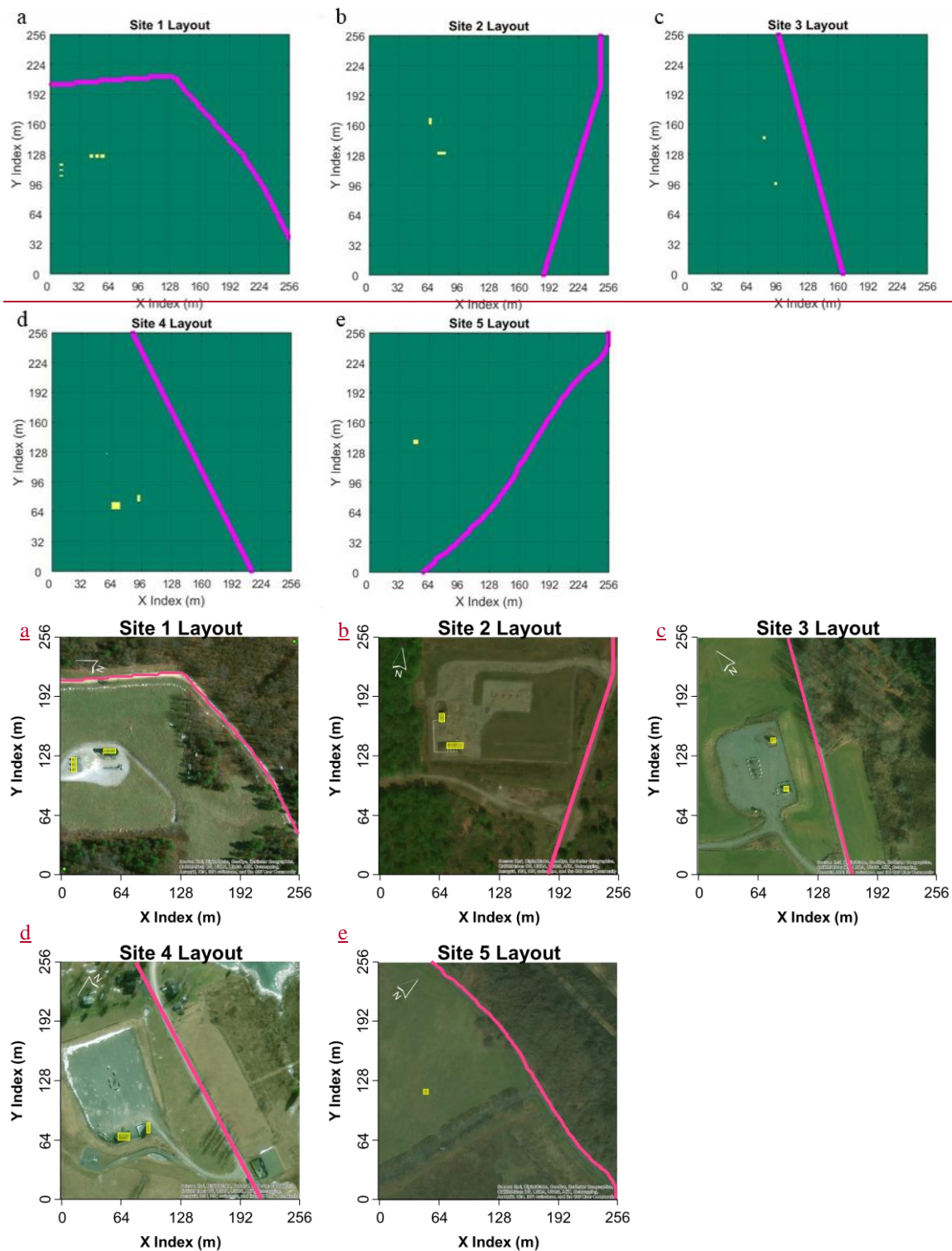
Figure 3. Site layouts showing emitting structures in yellow and the road in magenta for (a) Site 1, (b) Site 2, (c) Site 3, (d) Site 4 and (e) Site 5. ~~For visibility the roads are shown thicker than reality~~Wind is always entering from the domain from the left.
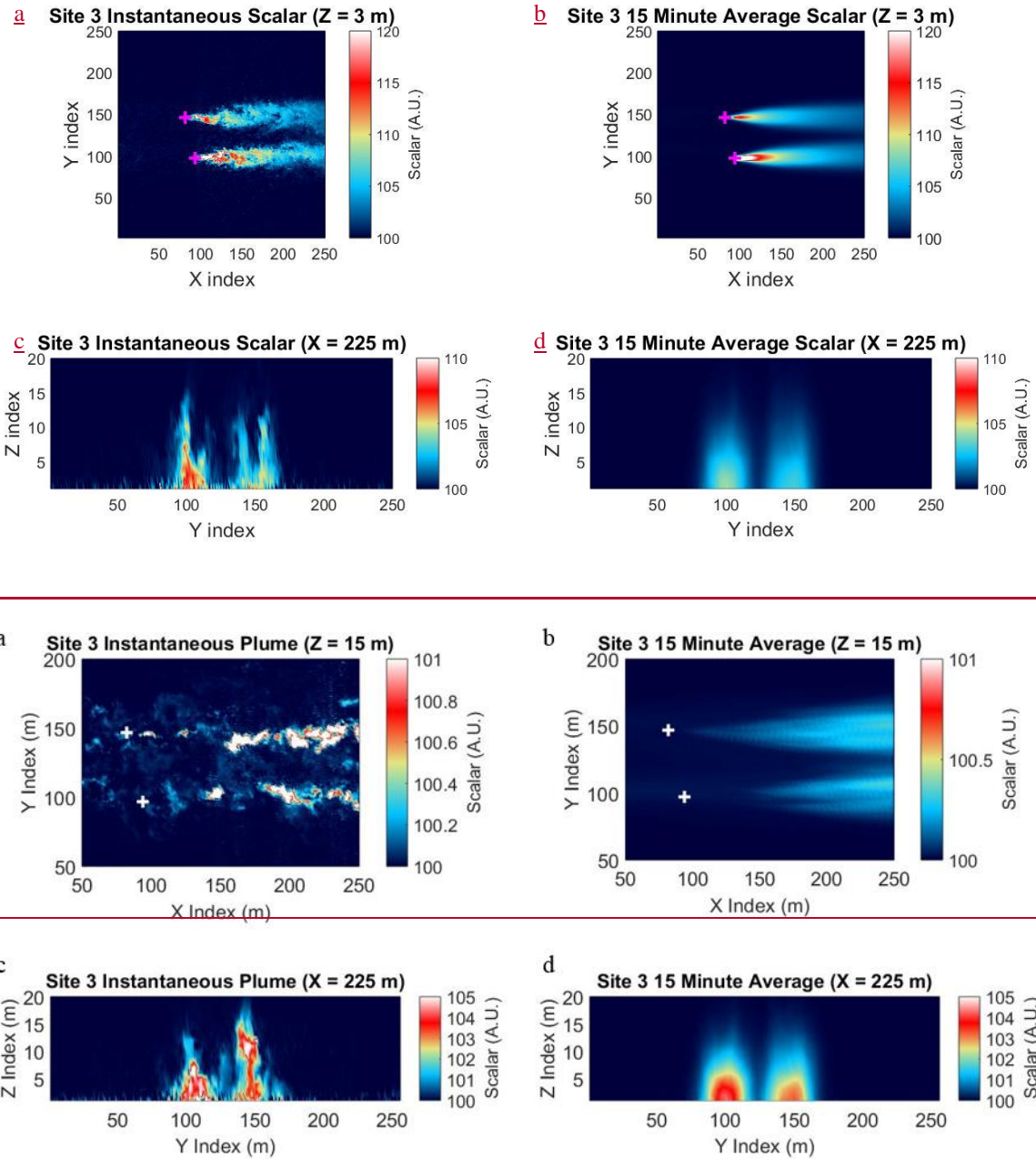
Figure 4. Comparison of instantaneous (a/c) and 15 min averaged (b/d) plume for Site 3~~.~~ from the LES. Panels a and b show scalar with arbitrary units (A.U.) in a x-y cross-~~ ~~section at an altitude of ~~15~~3 m and panels c and d show a x-z cross-~~ ~~section at a 225 m downwind distance. The release locations are shown with a ~~white~~magenta marker (+).
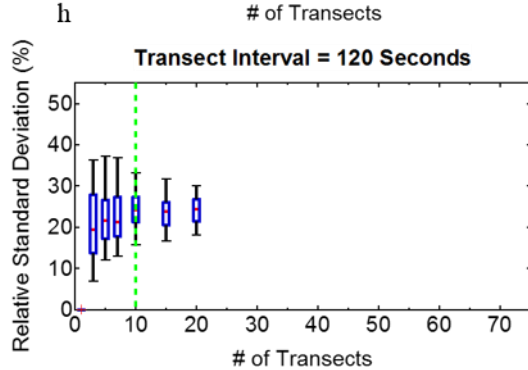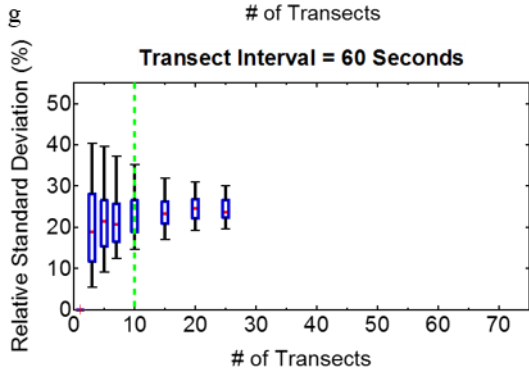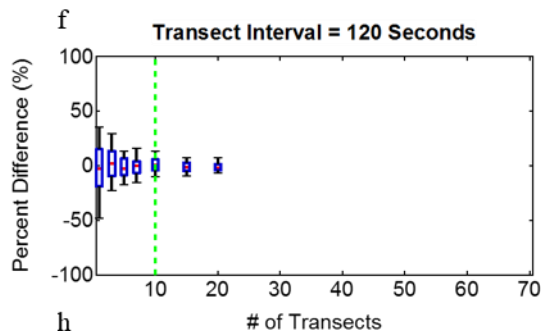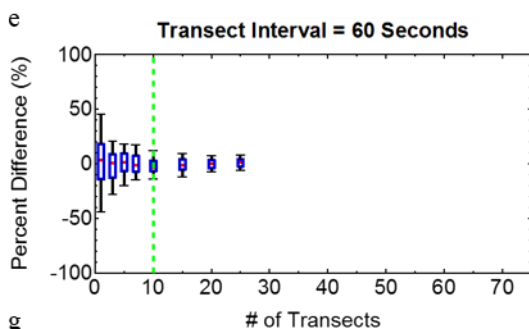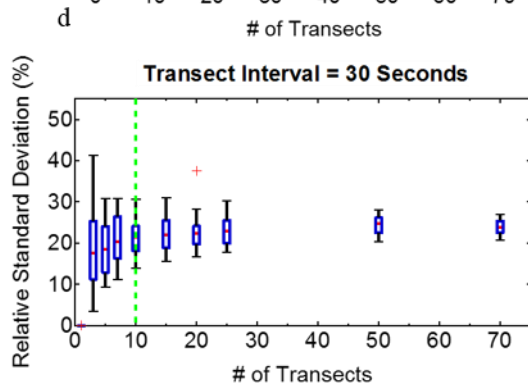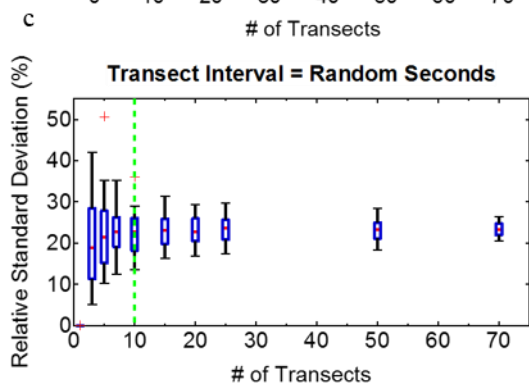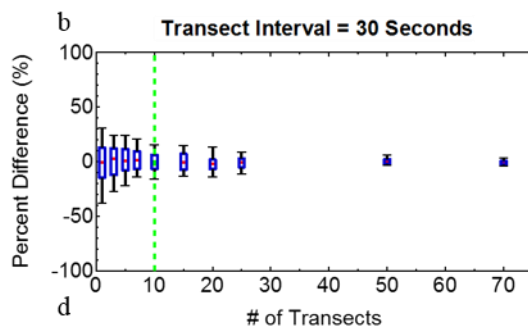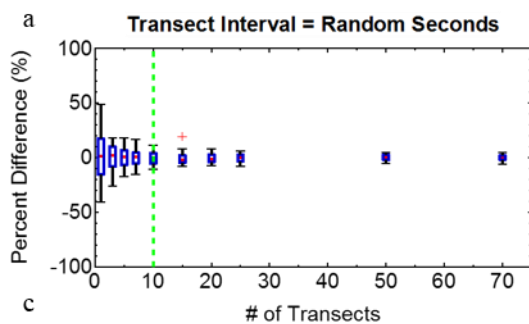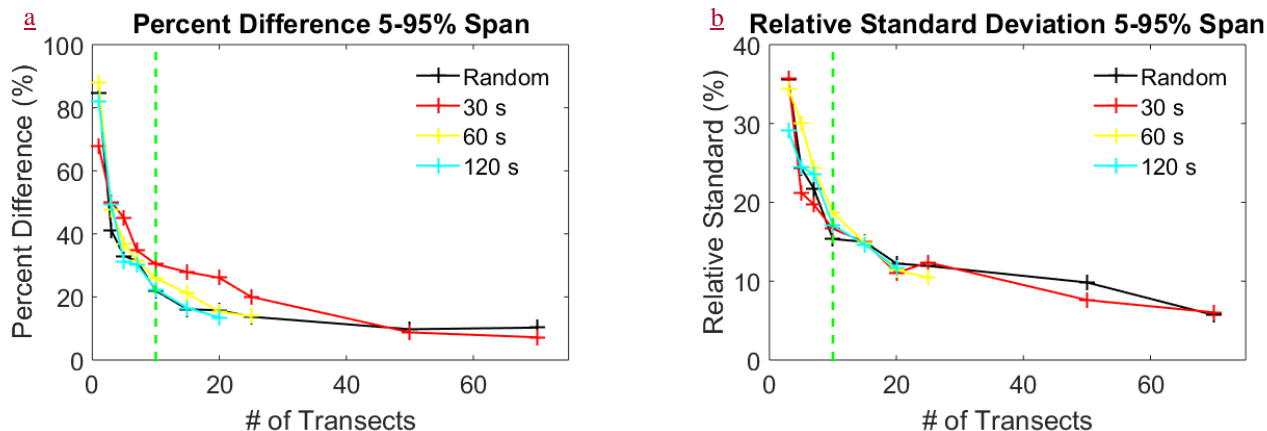
Figure 5. The 5-95% percent difference (a,b,e,f) between emission retrieval and known simulation emission rate and rsd (c,d,g,hb) of the emission retrieval using various amounts of transects and random (a,c), 30 second (b,d), 1 minute (e,g) or 2 minute (f,h) transect spacing. ~~Box and whiskers plots show the 50% percentile (red), 25 and 75% percentile (blue) and 10 and 90% percentile (black).~~ The recommended 10 transect criteria is shown in green.
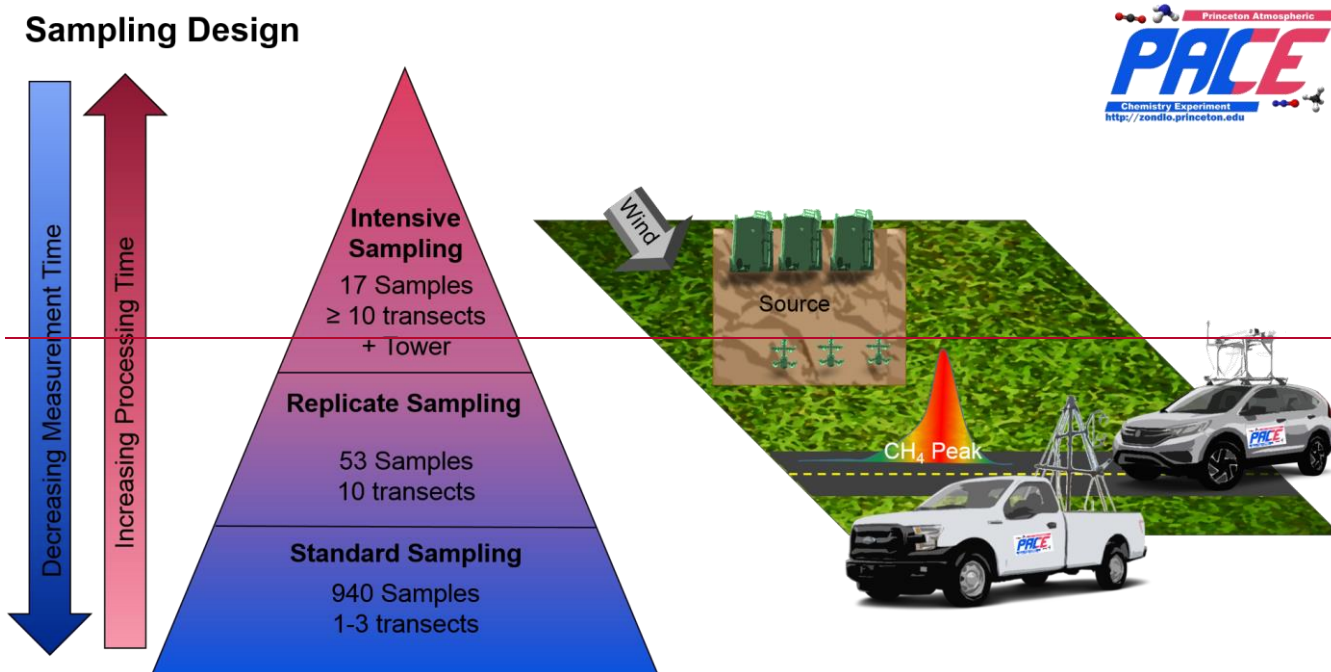
Hierarchical Sampling and Emission Rate Calculation Scheme
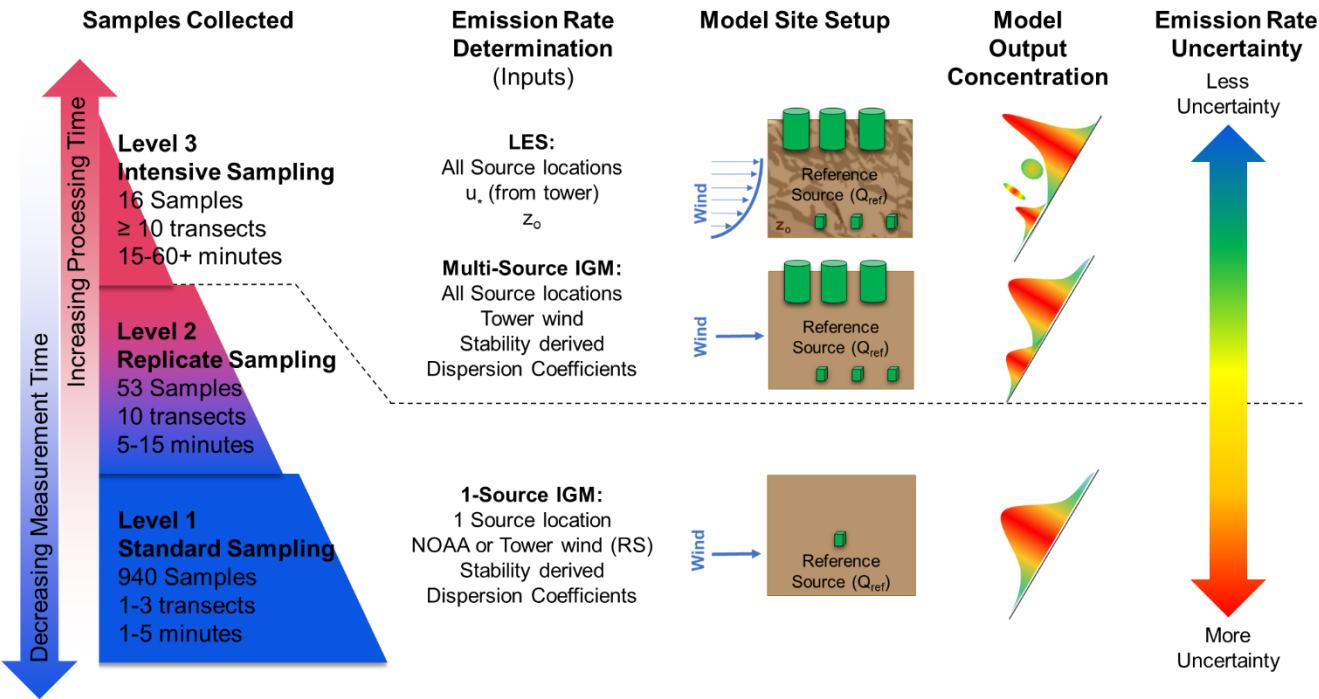
Figure 6. Finalized sampling and emission rate calculation strategy employed in this study showing actual measurements with increasing complexity and decreasing sample size.

Figure 7. ~~Finalized source strength determination strategy employed in this study showing~~ <u>as well</u> <u>as</u> models with increasing complexity and decreasing uncertainty. In this schematic u_* is friction velocity and z_o is terrain roughness length. <u>The dashed lines denote the pathways data typically followed. Standard and Replicate sampling always went through 1-Source IGM pathway. Note that 1-source IGM can also be calculated for Intensive Samples as is shown for comparison purposes in this work. Abbreviations: Large Eddy Simulation (LES), IGM (Inverse Gaussian Model).</u>

Site 3

a Raw Observations

b Centered Observations

Figure 87. **(a) Site 3 raw observations as indexed to the LES domain and (b) after they have been aligned with the peak of the LES plume. Waterfall plots of the same data for (a) raw observations and (b) centered observations.**

55
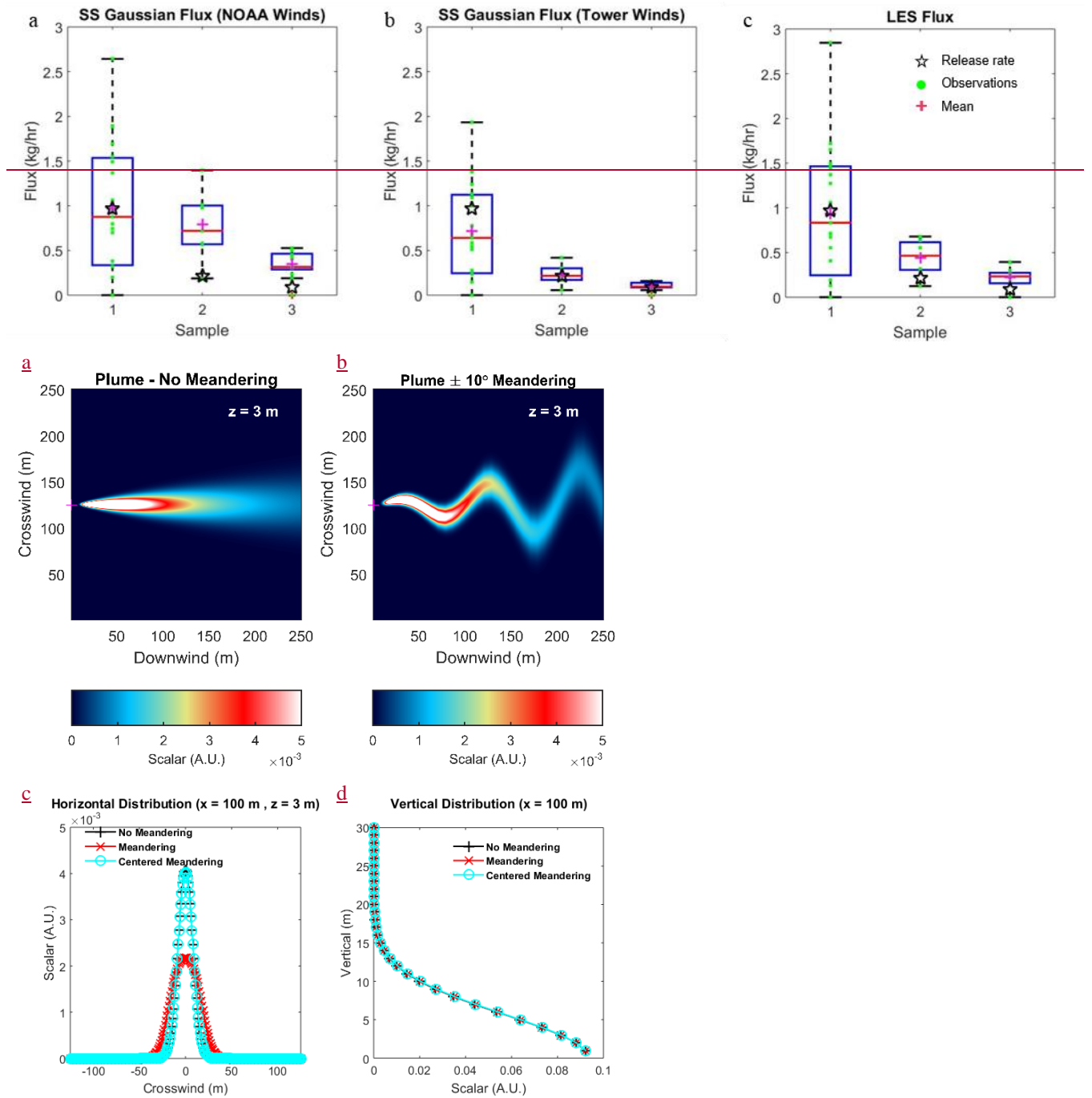
Figure 8. A theoretical site with an additional plume meandering scale added shown in a top-down view at 3 m of (a) Gaussian with no meandering, (b) instantaneous Gaussian with 10° meandering. The comparison of (c) horizontal

distributions of concentration and (d) vertical distributions of concentrations for three scenarios (no meandering, averaged meandering and centered meandering) are also shown. All units are arbitrary.
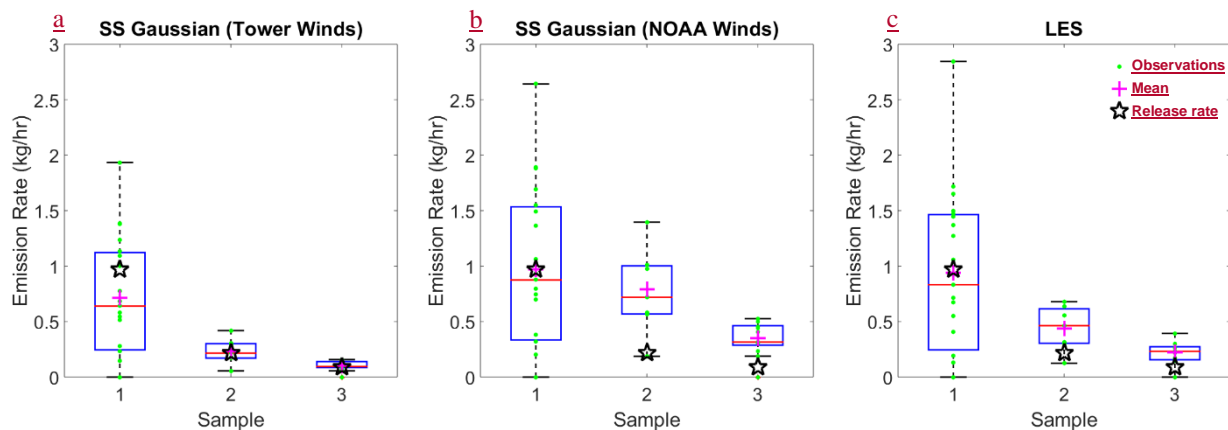
Figure 9. Results from three controlled release experiments for (a) the Gaussian approach using NOAA winds, (b) the Gaussian approach using tower measured winds and (c) from the LES. Box and whiskers plots show the 50% percentile (red), 25 and 75% percentile (blue) and minimum and maximum values (black). The mean is shown in magenta, green dots represent individual measurements and the black star is the actual release rate.

a

Release Rate (0.97 kg/hr)

b

Release Rate (0.22 kg/hr)

c

Release Rate (0.09 kg/hr)
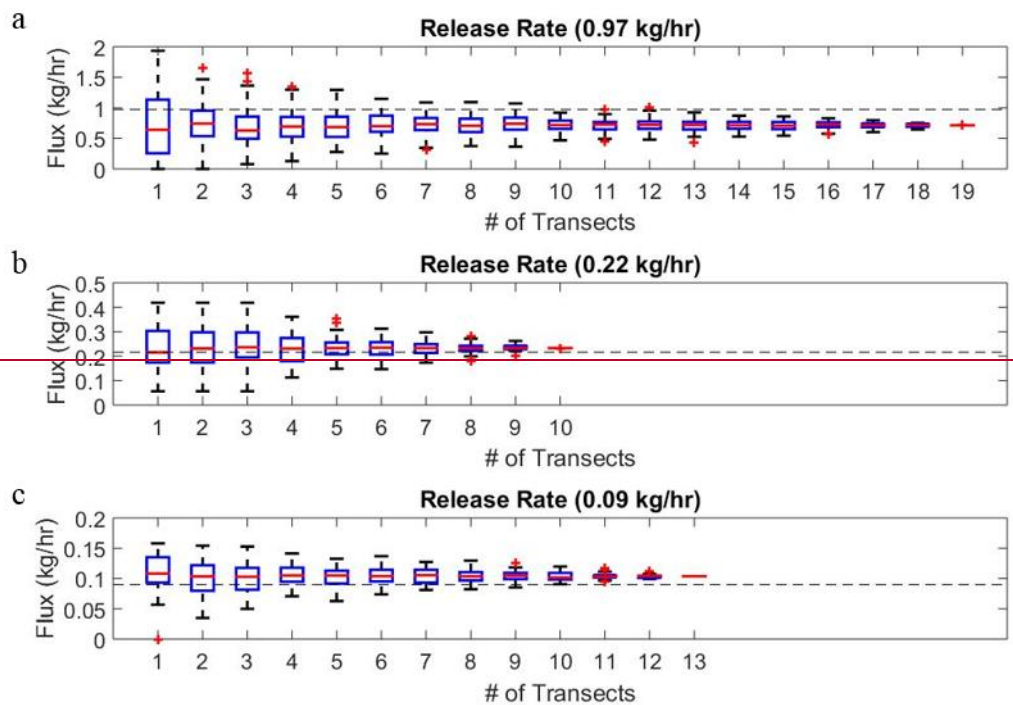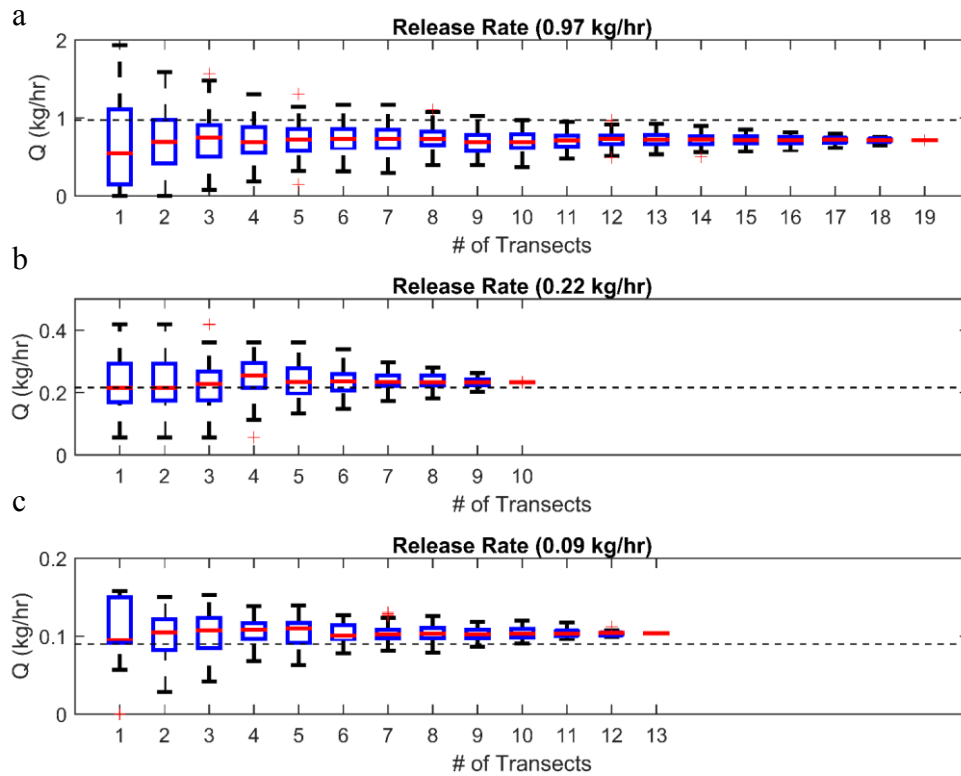
Figure 10. A comparison of the convergence of the plume rate inferred from the IGM by averaging randomly selected transects for (a) the 0.97 kg hr^{-1} release rate, (b) the 0.22 kg hr^{-1} release rate and (c) the 0.09 kg hr^{-1} release rate. The actual release rate is shown as a dashed black line. Box and whiskers plots show the 50% percentile (red), 25 and 75% percentile (blue) and minimum and maximum values (black). Outliers are shown in red.
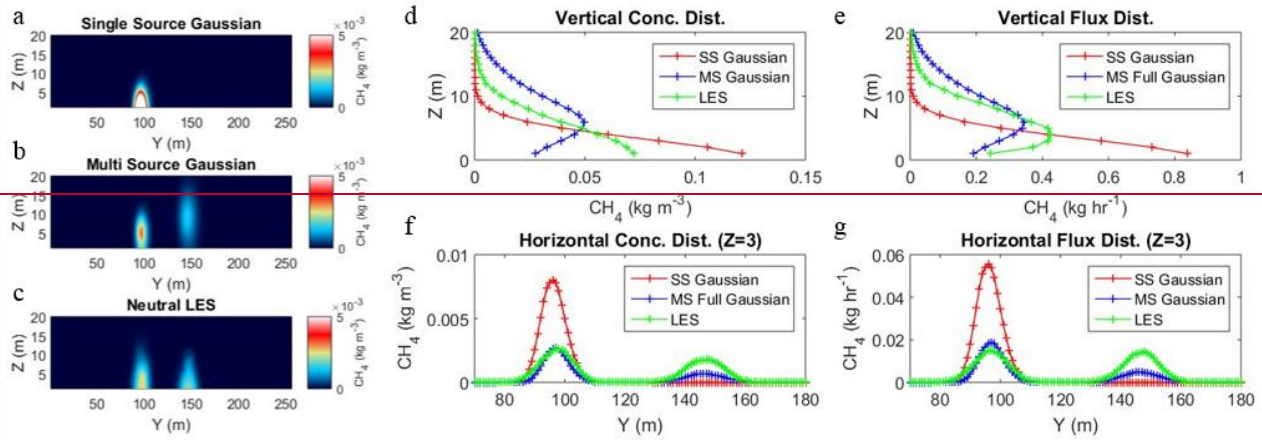
Site 3
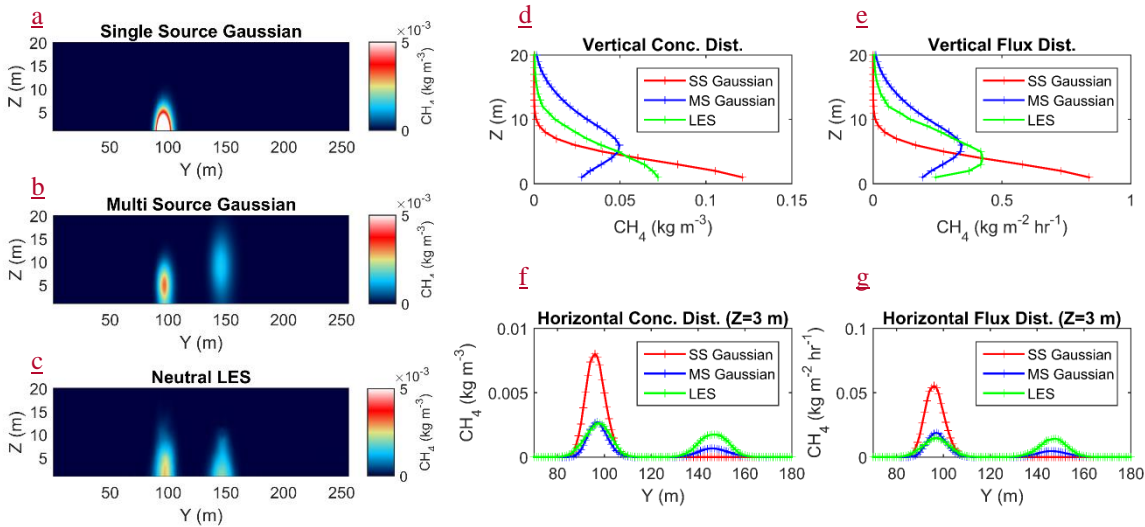
Figure 11. Comparison of three scenarios for Site 3 showing images through the downwind road plane (~30m downwind as shown in Fig. 3) of (a) single source Gaussian, (b) multi-source Gaussian and (c) averaged LES. The comparison of vertical distributions of (d) concentrations and (e) fluxes and of the horizontal distributions of (f) concentrations and (g) fluxes are also shown. The vertical LES flux corresponds to resolved fluxes only.

61
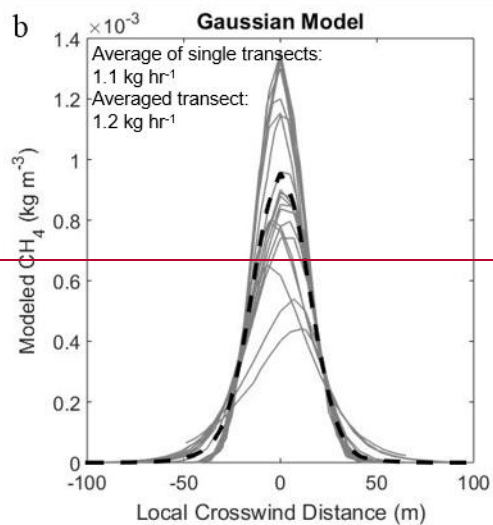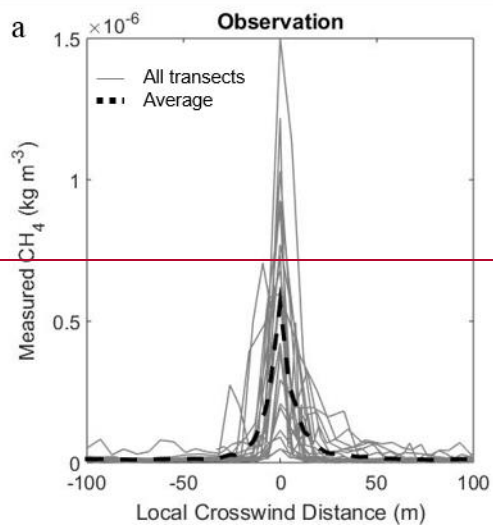
Figure 12. (a) Observations of CH₄ at Site 3 showing multiple downwind transects in gray and the averaged transect in black. (b) Gaussian outputs in gray along the same downwind transects and the averaged profile in black.

5

10

Supplemental

Table S1. Summary of LES site conditions and parameters.

Characteristic	Site 1	Site 2	Site 3	Site 4	Site 5*
Sample Date	7/15/2015	7/21/2015	6/21/2016	6/23/2016	8/5/2016
Time (EDT)	18:15-20:15	7:00-9:15	9:45-10:30	20:00-20:30	16:15-20:00
Wind Speed (m s^{-1})	1.30 ± 0.64	1.30 ± 0.70	1.9 ± 2.2	0.62 ± 0.27	1.70 ± 0.89
Wind Direction	325 ± 101	252 ± 40	276 ± 50	40 ± 47	218 ± 28
Friction Velocity (m s^{-1})	0.26	0.29	0.48	0.18	0.10-0.21
Distance from Source (m)	176	154	29	49	110
Elevation gain (m)	3.5	1.5	2	2	2
Simulation Time (min)	59.2	33.33	33.33	33.33	30
Simulation Warm up Time (min)	10	5	5	5	5
Simulation x-dimension (m) (x resolution)	288 (2)	256 (1)	256 (1)	256 (1)	256 (1)
Simulation y-dimension (m) (y resolution)	256 (2)	256 (1)	256 (1)	256 (1)	256 (1)
Simulation z-dimension (m) (z resolution)	100 (1)	100 (1)	100 (1)	100 (1)	33.33 (0.2222)

* Controlled release site.

Figure S1. Controlled release site schematic.
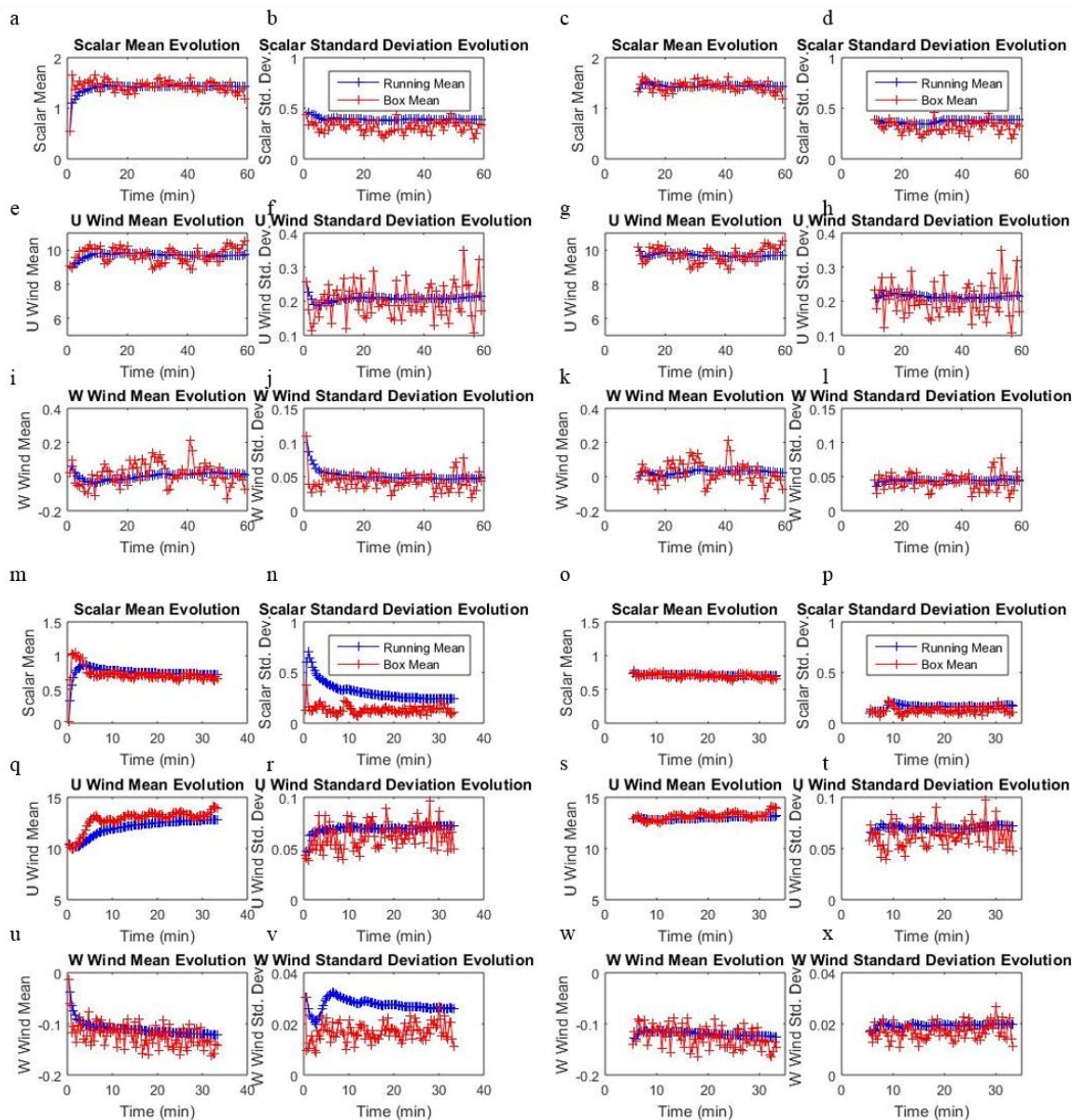
Figure S2. Comparison of box-means or standard deviations (red) and running means or standard deviations (blue) for (a-l) Site 1 and (m-x) Site 2. Left-hand panels show both sites from start-up and right-hand panels are screened to show results after a steady-state has been achieved. The running mean in the screened results starts at the determined onset of steady state.

65

Site 1
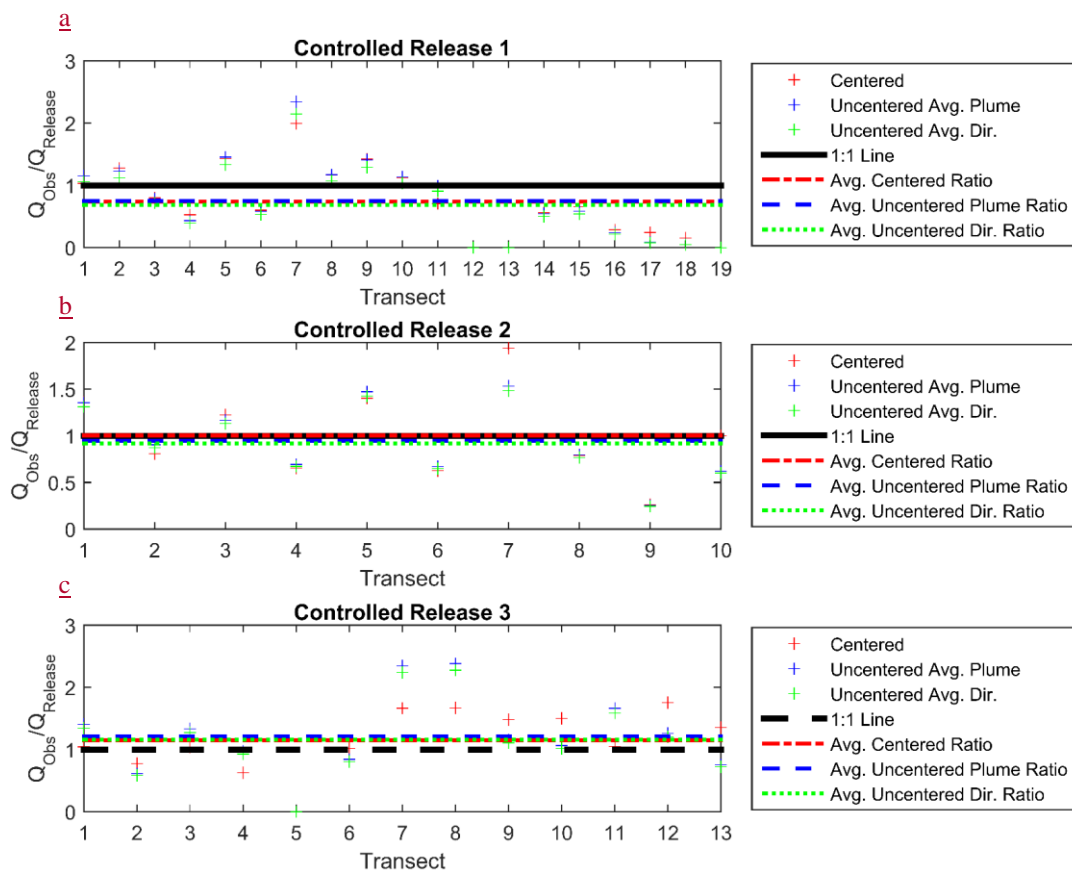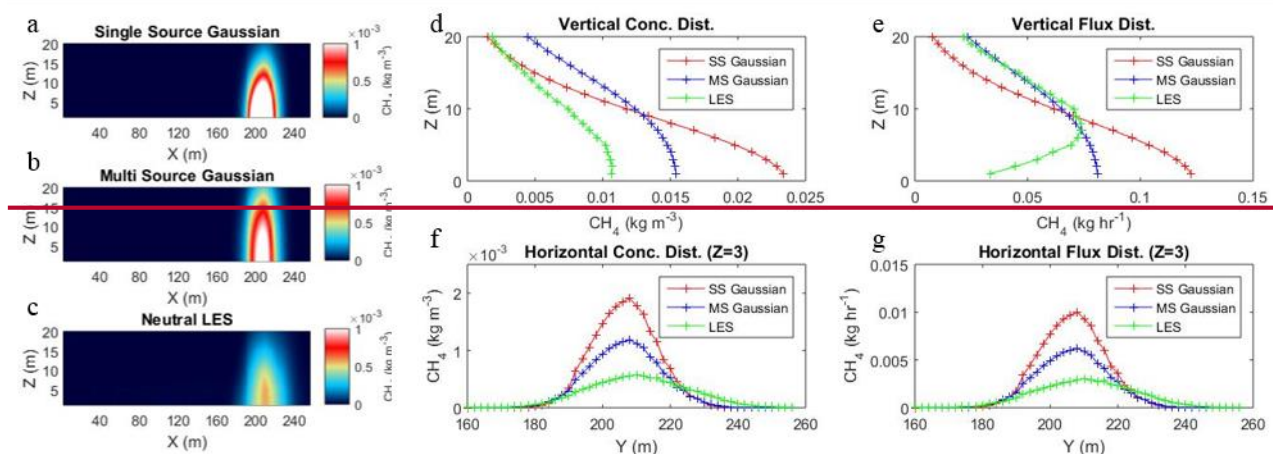
Figure S3. A comparison of release experiments 1-3 (a-c) results using the Gaussian aligned to the observation peaks (Centered) and the average Gaussian (Uncentered Avg. Plume) and the single Gaussian in the average wind direction (Uncentered Avg. Dir.).
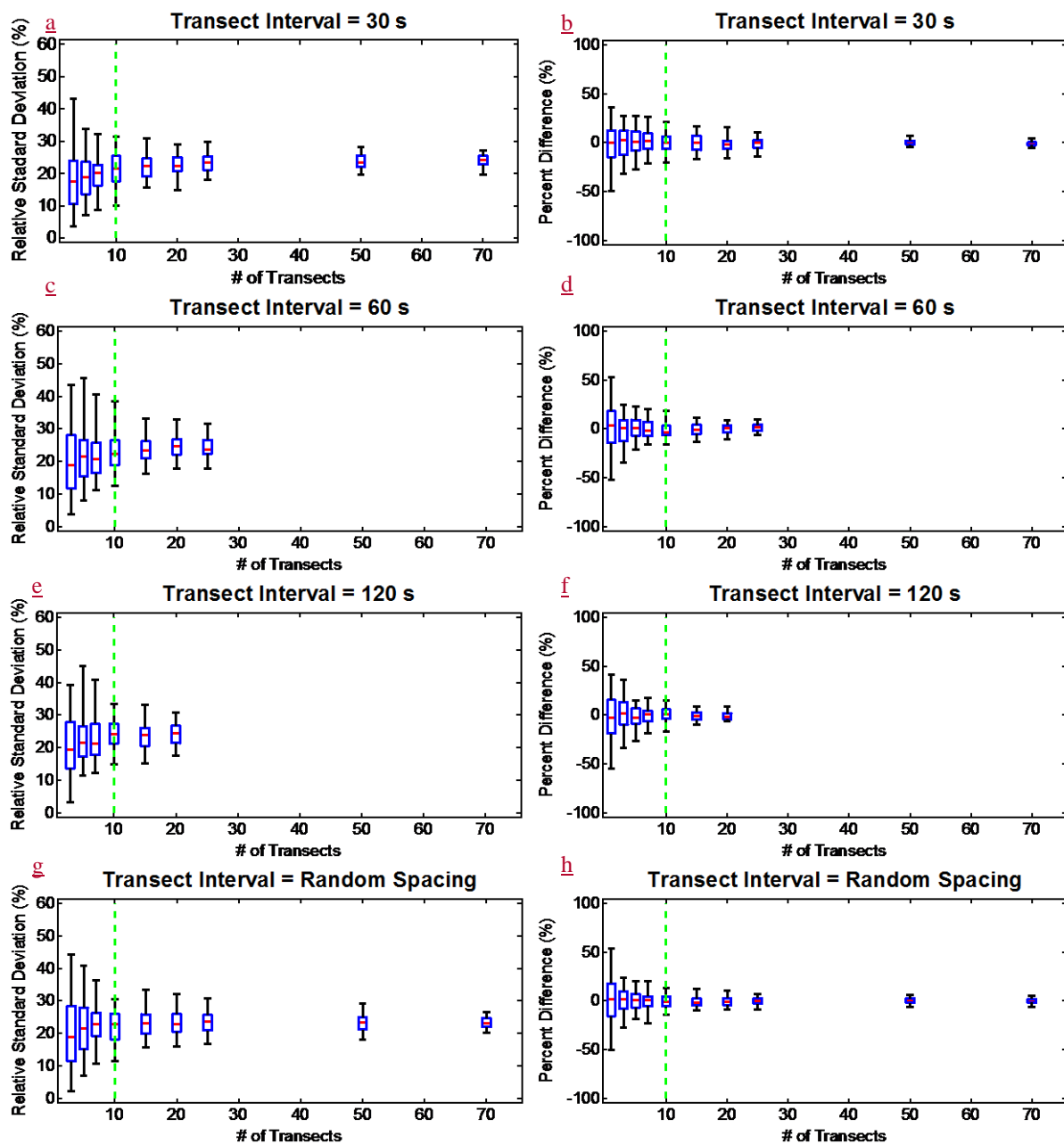
Figure S4. The rsd (a,c,e,g) of the emission retrieval using and the percent difference (b,d,f,h) between emission retrieval and known simulation emission rate using various amounts of transects and 30 second (a,b), 1 minute (c,d), 2 minute (e,f) or random (g,h) transect spacing. Box and whiskers plots show the 50th percentile (red), 25th and 75th percentile (blue) and 2.5 and 97.5 percentile (black). The recommended 10 transect criteria is shown in green.

68
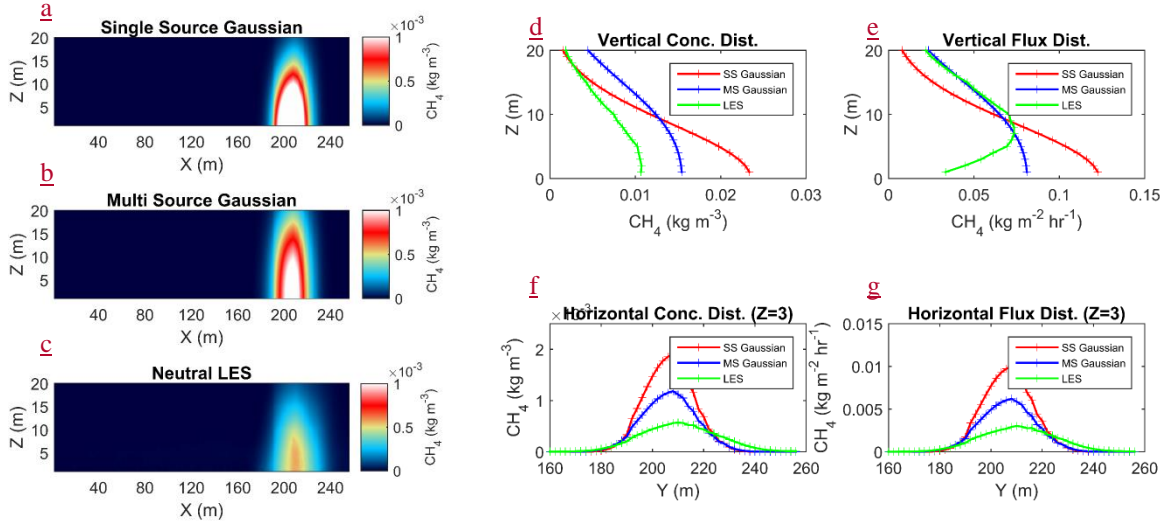
Figure S5. Comparison of three scenarios for Site 1 showing images of (a) single Gaussian, (b) multi-source Gaussian and (c) averaged LES. The comparison of vertical distributions (d) concentrations and (e) fluxes, and horizontal distributions of (f) concentrations and (g) fluxes are also shown.

Site 2
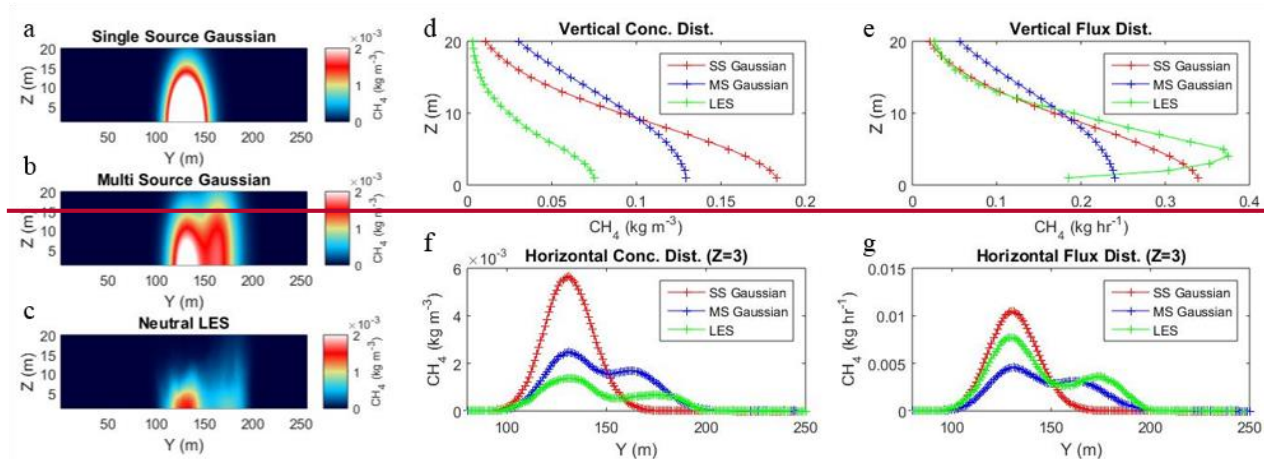
Site 2
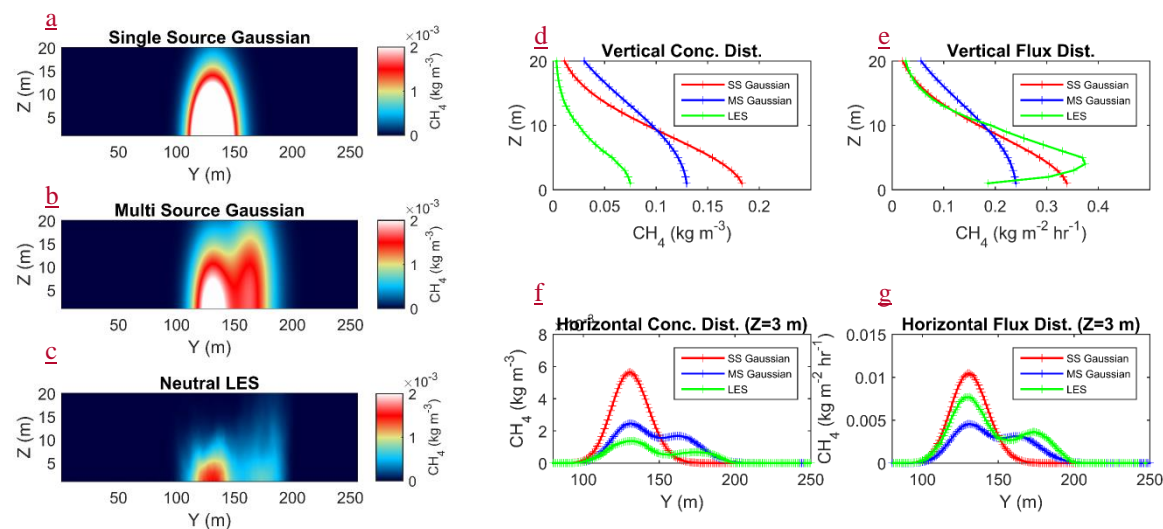
Figure ~~S4~~S6. Comparison of three scenarios for Site 2 showing images of (a) single Gaussian, (b) multi-source Gaussian and (c) averaged LES. The comparison of vertical distributions of (d) concentrations and (e) fluxes, and horizontal distributions of (f) concentrations and (g) fluxes are also shown.
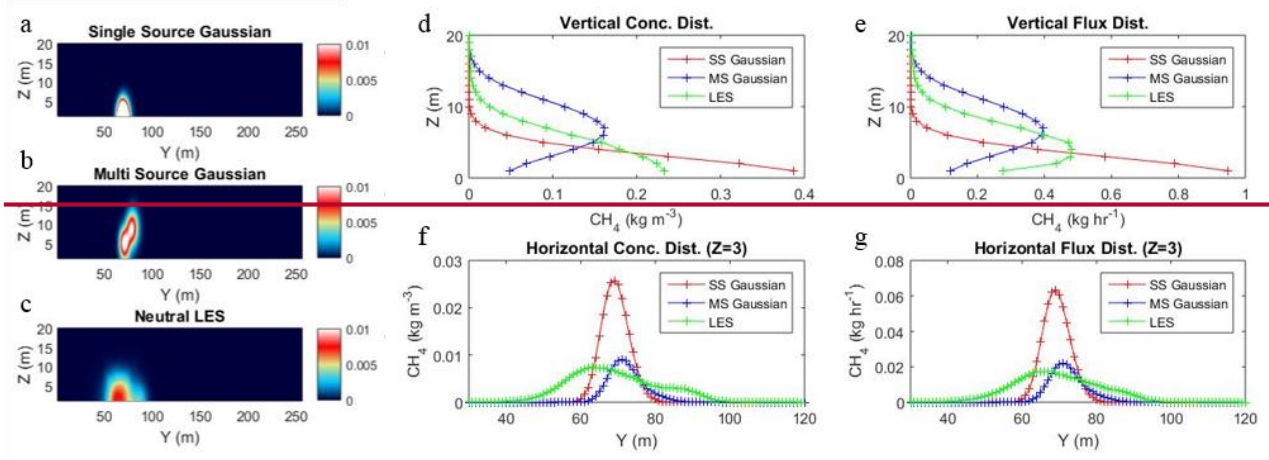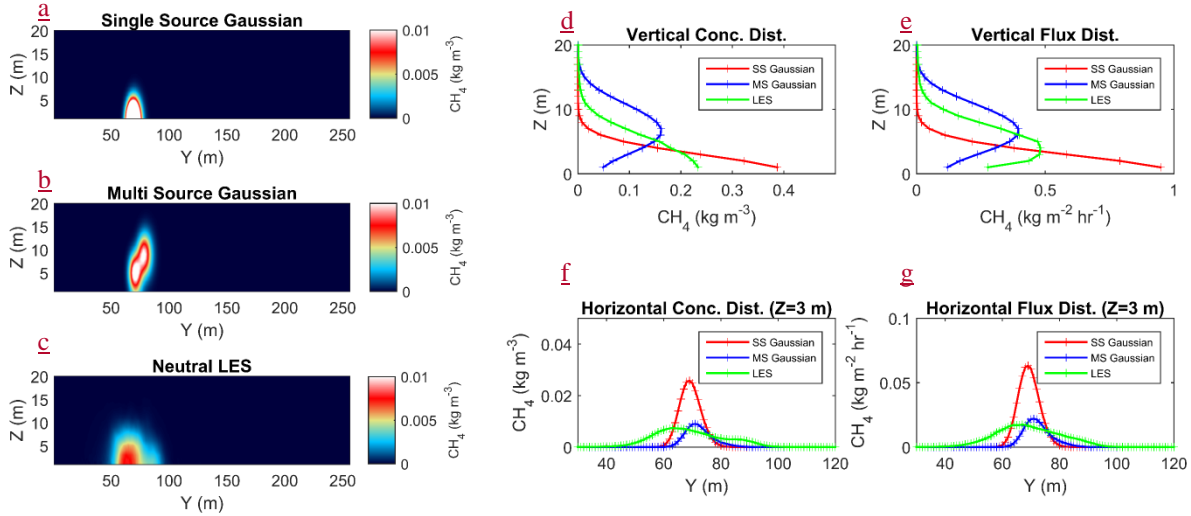
Figure ~~S5~~S7. Comparison of three scenarios for Site 4 showing images of (a) single Gaussian, (b) multi-source Gaussian and (c) averaged LES. The comparison of vertical distributions of (d) concentrations and (e) fluxes, and horizontal distributions of (f) concentrations and (g) fluxes are also shown.

71
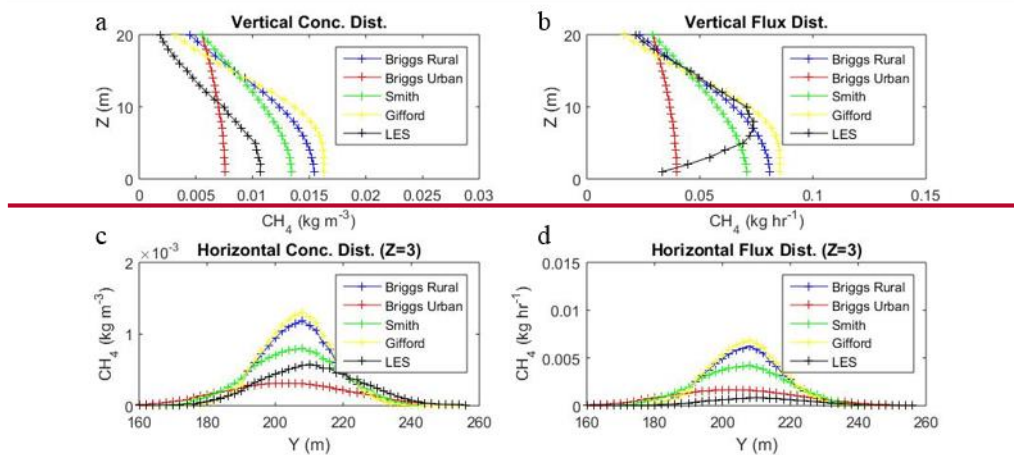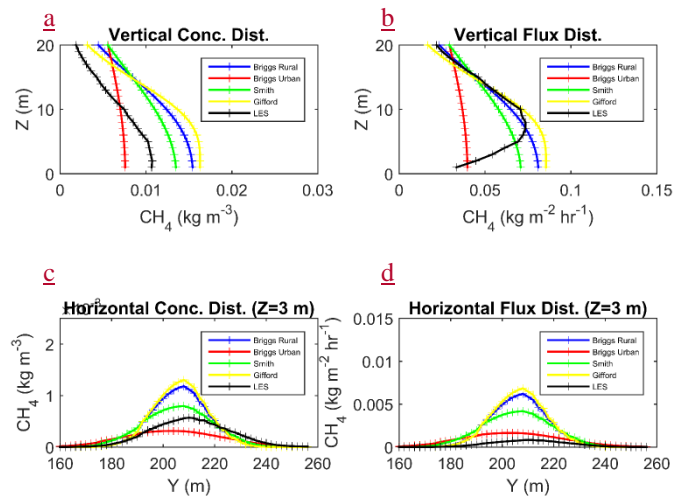
Site 1

Figure S6S8. Comparison of scenarios using different diffusion models for Site 1 showing images of the comparison of vertical distributions of (a) concentration, (b) flux and of the horizontal distributions of (c) concentration and (d) flux.

72

Site 2

a — Vertical Conc. Dist.
b — Vertical Flux Dist.
c — Horizontal Conc. Dist. (Z=3)
d — Horizontal Flux Dist. (Z=3)

Site 2

a — Vertical Conc. Dist.
b — Vertical Flux Dist.
c — Horizontal Conc. Dist. (Z=3 m)
d — Horizontal Flux Dist. (Z=3 m)
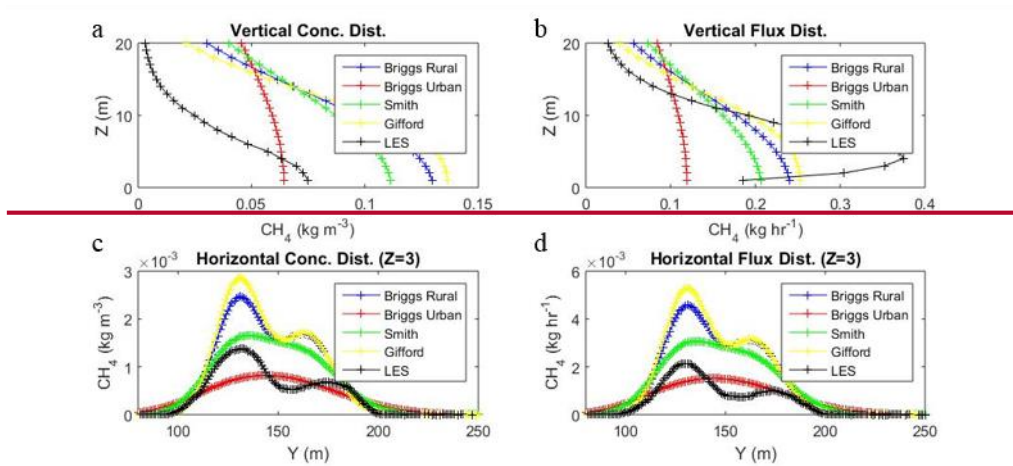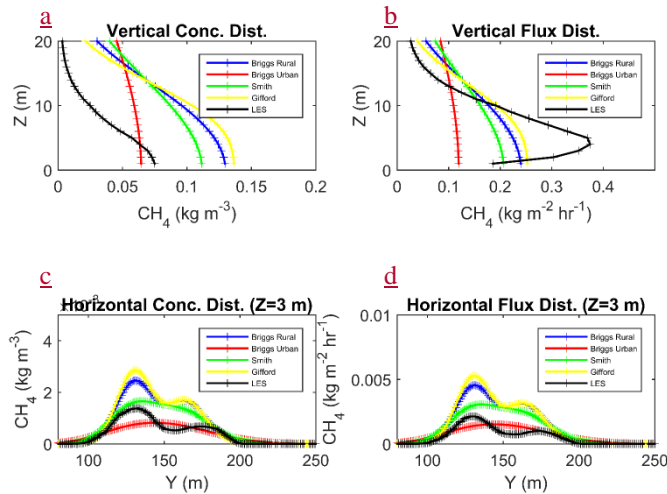
Figure S7S9. Comparison of scenarios using different diffusion models for Site 2 showing images of the comparison of vertical distributions of (a) concentration, (b) flux and of the horizontal distributions of (c) concentration and (d) flux.
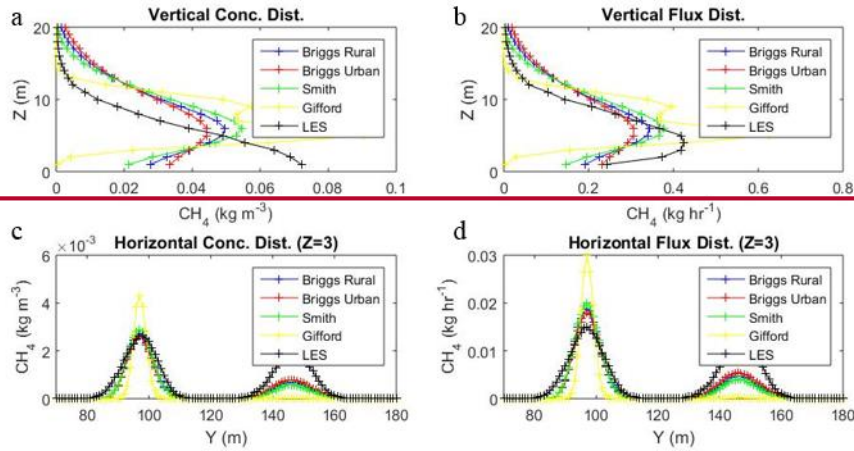
Site 3
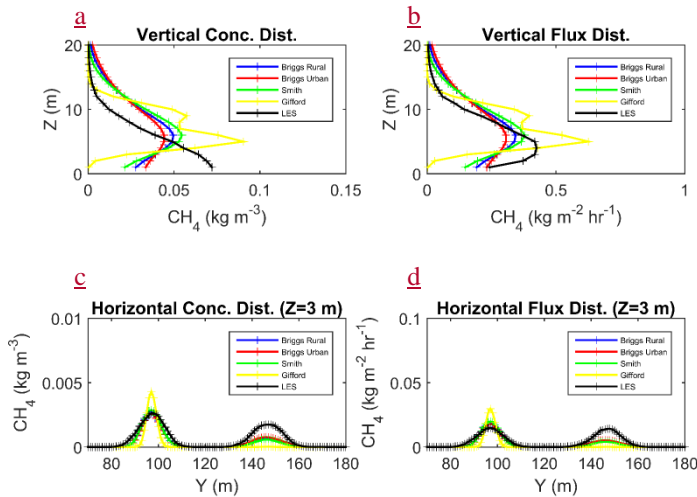
5   **Figure S8S10. Comparison of scenarios using different diffusion models for Site 3 showing images of the comparison of vertical distributions of (a) concentration, (b) flux and of the horizontal distributions of (c) concentration and (d) flux.**

74
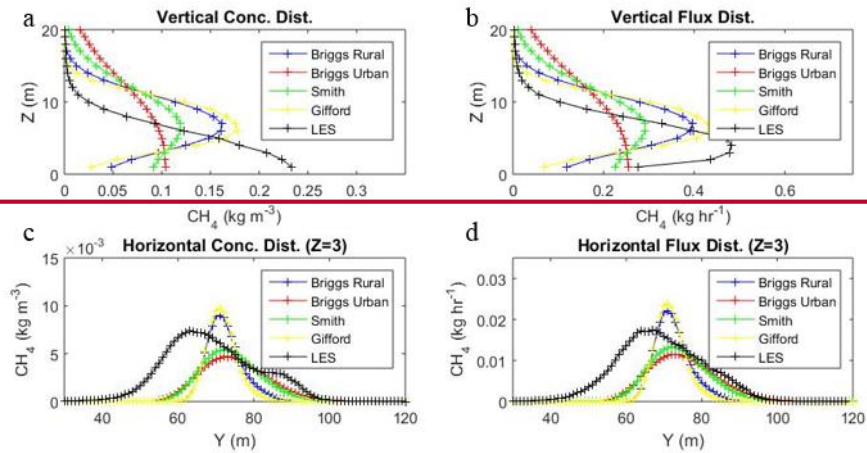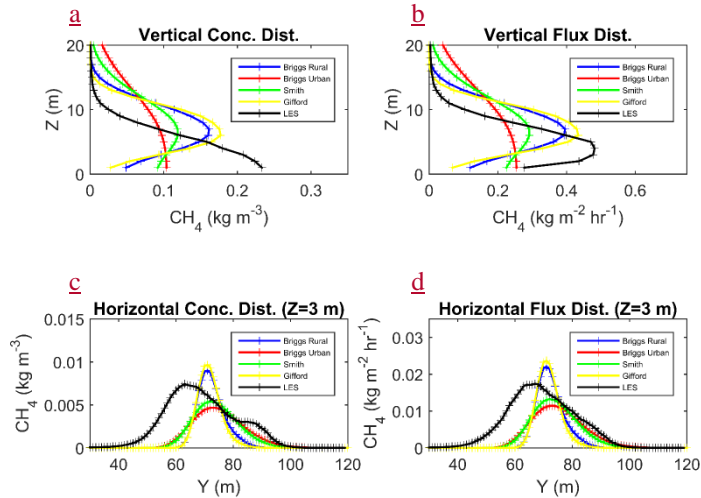
Site 4

Figure S9S11. Comparison of scenarios using different diffusion models for Site 4 showing images of the comparison of vertical distributions of (a) concentration, (b) flux and of the horizontal distributions of (c) concentration and (d) flux.

75

a

Source Location Uncertainty Effect

b
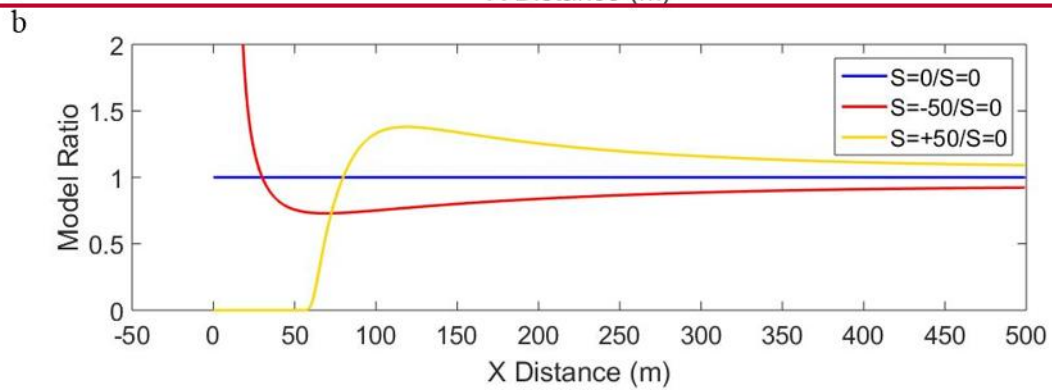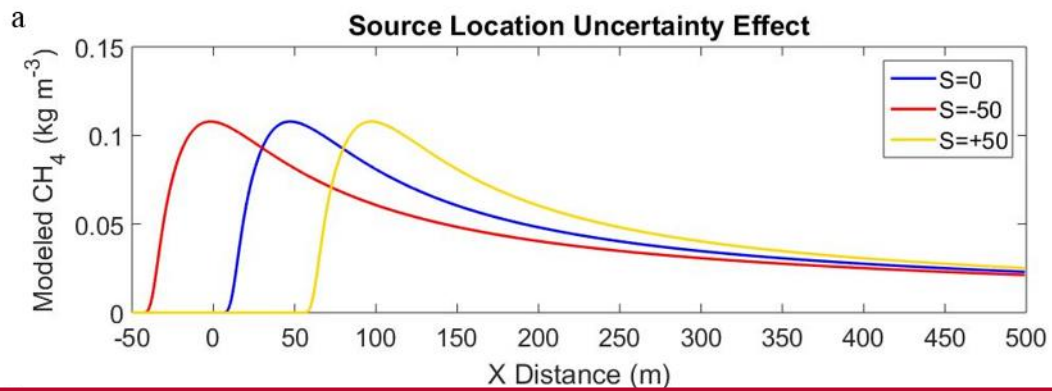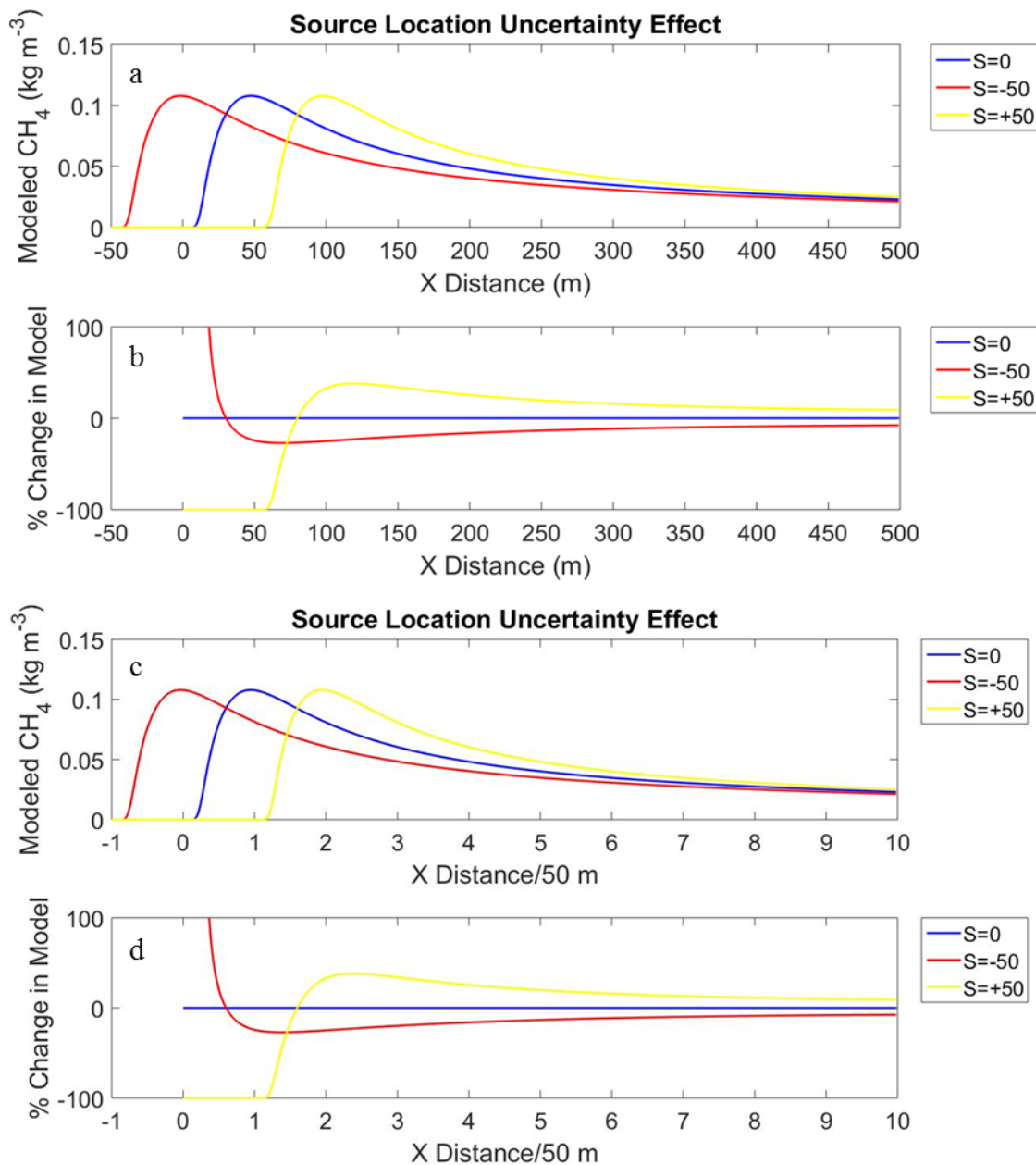
Figure S10S12. (a) Modeled CH4 at three source locations at different _downwind_ x positions assuming 3 m receptor height, 1 m source height, neutral stability and 1.5 m s^{-1} wind speed. **(b)** ~~Ratio between the sum~~Change in ~~y and distance x~~% of the modeled CH4 in each of the scenarios and base scenario (S=0). _(c) The same as panel (a), normalized by the perturbation distance of 50m. (d) The same as panel (b), normalized by the perturbation distance of 50 m._
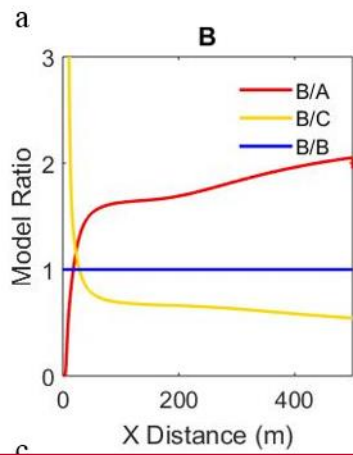
Source Height Uncertainty Effect
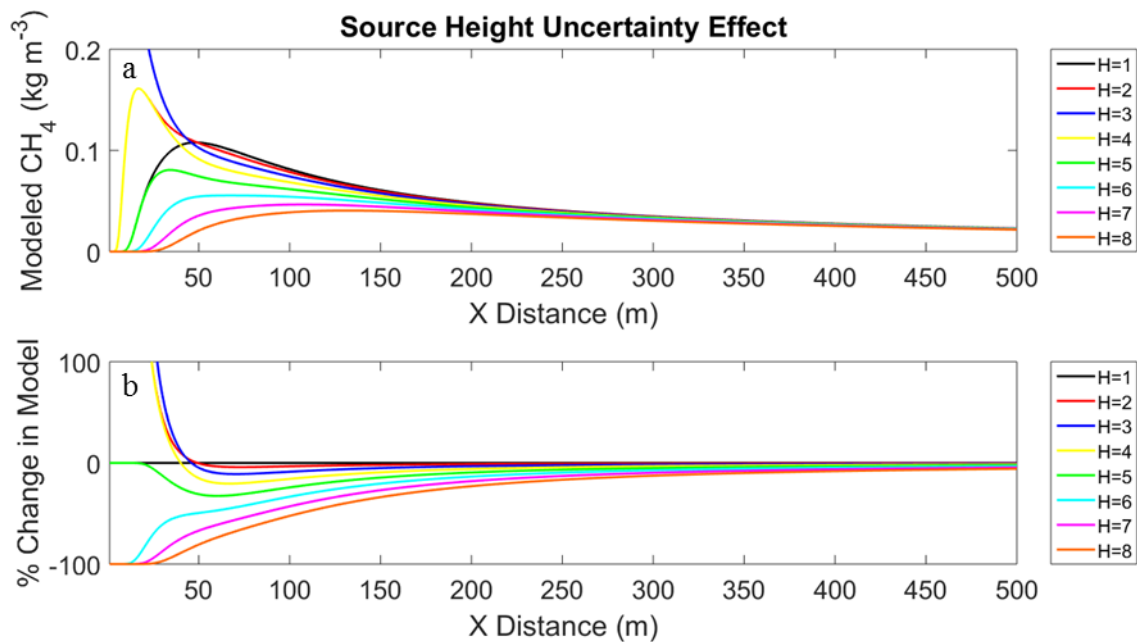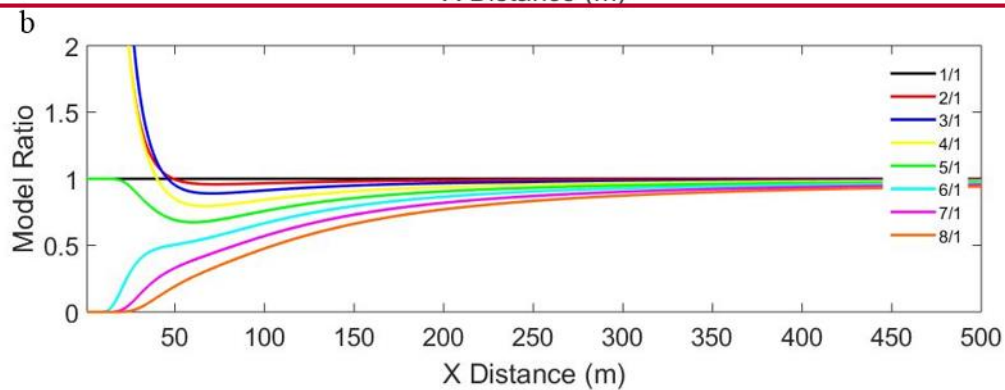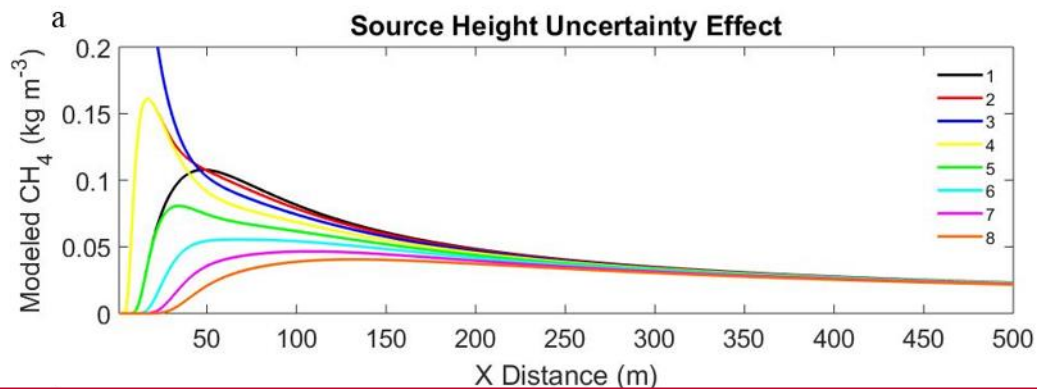
Figure S11S13. (a) Modeled CH4 using 8 source heights assuming 3 m receptor height, neutral stability, and 1.5 m s⁻¹ wind speed. (b) RatioChange in % between the summodeled CH4 for each of the scenarios and the base scenario (h=1 m).

5
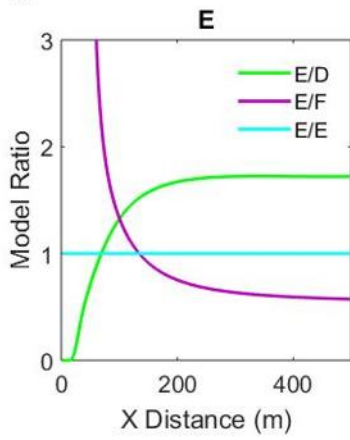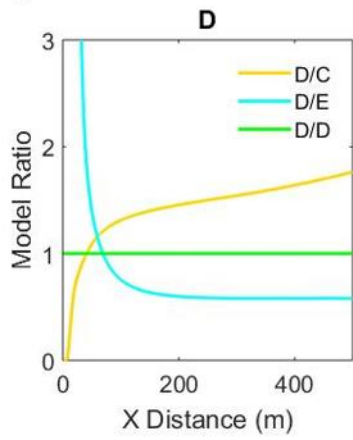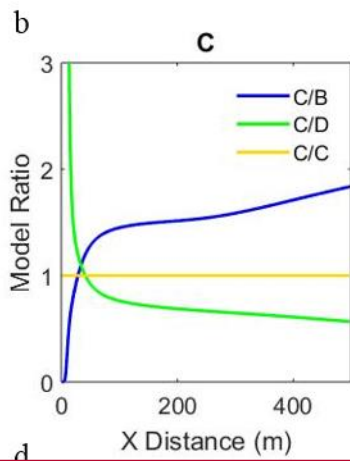
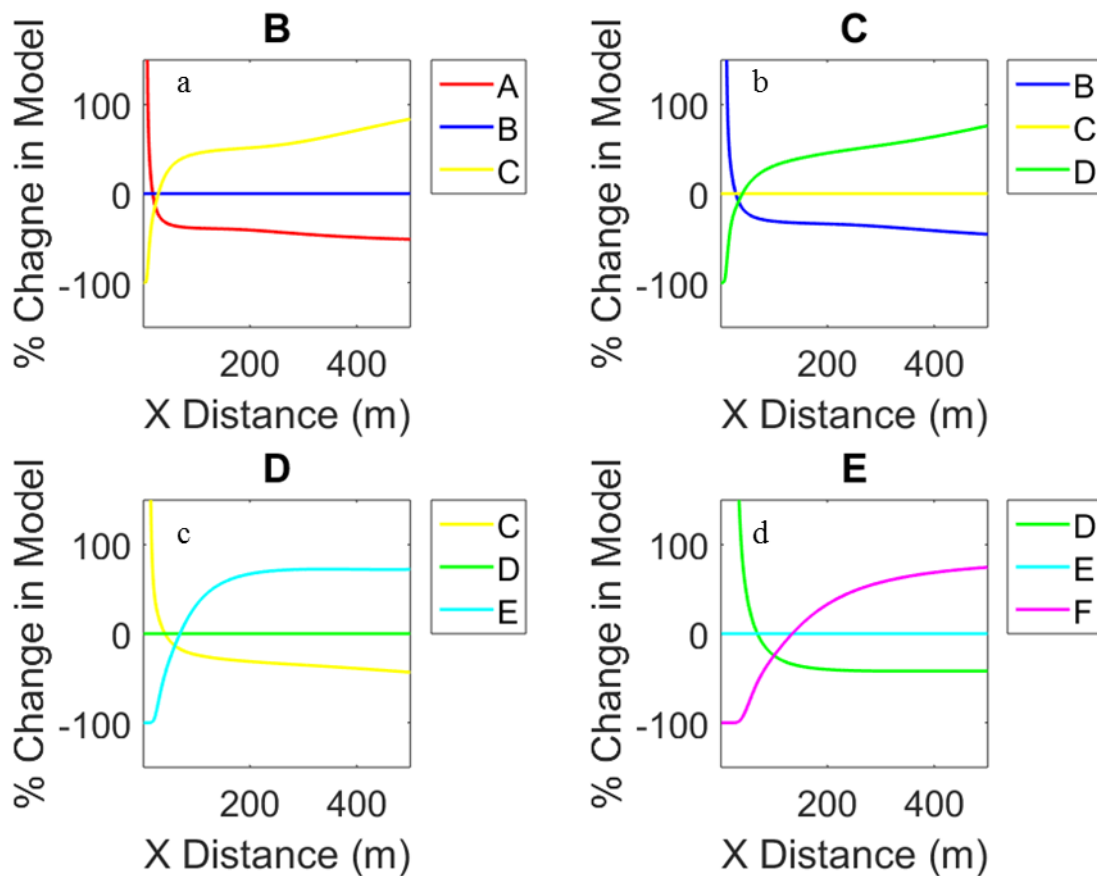**Figure ~~S12. Comparison~~S14. Change in percent between ~~Gaussian model sums and~~ modeled CH4 between +/- one stability class and base scenario for (a) Class B, (b) Class C, (c) Class D, and (d) Class E assuming 3 m receptor height, 1 m source height, and 1.5 m s^{-1} wind speed.**
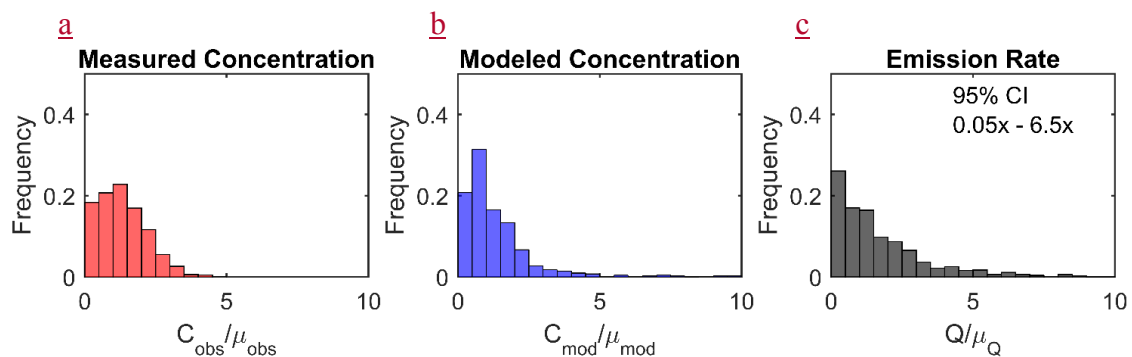
Figure S15. Monte Carlo simulations using 1,000 replicates for the standard sampling, 1 transect case showing (a) normalized observation distribution, (b) normalized model distribution and (c) normalized emission rate distribution.