Responses to Anonymous Referee #1

We thank the reviewer for the meticulous review. The comments have helped to refine and clarify the manuscript. All comments are addressed below and changes in text, figures, captions or tables have been included as appropriate. 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 #1 Comments

Major issues.

1. Basic issues concerning the methods are not explained, thus many of the most important results cannot be clearly interpreted. I have included specific comments for many figures, tables, and text in my detailed comments. I mention what I consider to be the most important overall points here.

a. **Table 5 is arguably the most important product of the manuscript, but the basic elements that are used to build this table are never clearly articulated. The manuscript should be reorganized to clearly explain the basis for the uncertainties being imposed in the inputs to the Gaussian plume emissions estimates, and the uncertainties in these inputs should be clearly stated. This is true with a limited number of inputs in this table (instrument precision, stability class), though even for one of these inputs, the source of the uncertainty in the inputs to the model is not clear (e.g. why is one class the assumed bound of uncertainty?, what time resolution is associated with 5 ppb precision in the LiCor 7700?). Atmospheric variability, the most important issue according to the table, is never defined. The main finding of the manuscript must be clearly defined to be useful and interpretable.**
   Various clarifications have been made throughout the text that we will reference in response to later specific questions that we believe help clarify the methodology. For instance stability classification is discussed in Sect. 5.1 (p. 15-16). The time resolution and uncertainty of the LI-COR are now included in Table 6 (previously Table 5). While atmospheric variability was discussed in Sect. 3.1, we have added an additional clarification on p.4 lines 24-25.
   "Insufficient averaging would result in random errors that we refer to in this paper as the uncertainty related to 'atmospheric variability'."
   Additional clarification of this concept has also been made on p. 12 lines 2-6.
   "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)."
   We have also added additional discussion in Sect. 5.2 to facilitate comprehension of this concept as we have specifically used it and at this important section of the work on p. 20 lines 22-23.
   "The largest contributor to total uncertainty is atmospheric variability, the random error induced by insufficient averaging of the turbulent instantaneous plume."
b. **The method for determining single vs. multiple sources, or the method for evaluating the number and spatial distribution of multiple sources, is never explained.**

Additional discussion of how sources were determined has been added to Sect. 4.1. on p.13 line 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."*

c. The method used to remove background concentrations from the field measurements is critical, and is never discussed. This is a major source of uncertainty in plume dispersion estimates, and is not included in the results. This must be addressed.

We have added a line to Table 6 to indicate the expected contribution of the background selection. In addition, we have added the following text to Sect. 5.1. on p. 19 lines 10-19.

"*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."*

2. The averaging applied to the downwind transects is arguably inappropriate and suggests a fundamental misuse of the Gaussian plume model. The Gaussian plume model describes the ensemble average of atmospheric concentrations downwind of a point source. This includes the fact that the instantaneous plume will meander over the course of time due to atmospheric turbulence. The model is based on the assumption of a separation of scales between atmospheric turbulence, and mesoscale or synoptic scale atmospheric flow variations. The authors have chosen, however, to align each measurement downwind of their point sources by the peak concentration measured on each transect. The wind direction, therefore, changes with every transect. The mean wind in the Gaussian plume model should be the average wind across the time span of all of the downwind transects (which are conducted within an hour, over which the atmospheric dispersion conditions are assumed to be steady-state), and should not vary every few minutes with each transect. This approach to averaging would create a significantly broader observed plume. It is not clear to me how this might alter the estimated emissions, but this is a misapplication of the plume dispersion model that should be corrected. If the dispersion conditions within the time span of the downwind transects are not steady, this must be explicitly identified and treated in the analyses.

We thank the reviewer for this comment and the opportunity to better defend and define our work. We understand and agree the definition of 'meandering' which has been used in different ways in the literature has not been well defined for this work and the centering approaches have not been validated. For expediency we break the reviewer's comment into three points:

1) The definition of meandering in the context of this work and whether it is included in Gaussian models. Related to this is the choice to integrate to calculate emissions.
2) The centering of the Gaussian on individual transects to calculate emission rates for the bulk of our data.
3) The centering of the observations to the LES output for the 5 specific cases in this work.

Our responses are provided in this order below.

1) A point of clarification about the Gaussian plume must be made. As an analytical expression of plume diffusion, the Gaussian model will only include plume meandering in its output if the contribution of meandering scales is included in the diffusivity constants employed. This is related to Referee 2's point citing Fritz et al. (2005) discussion of the proper averaging intervals to produce plume cross sections that match the predicted model. They observed timescales ranging from a few minutes to hours meaning the model could over or under predict the cross section observed by the data. This is actually the physical basis for the integration used to calculate emission rates. The scales of 'meandering' as we define them pertain to scales outside the range the LES can include (>100m). However, these are not scales that do not contribute to vertical diffusion. While this would change the Gaussian model, the advantage of the IGM approach is that σ_y and σ_z are independent :

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

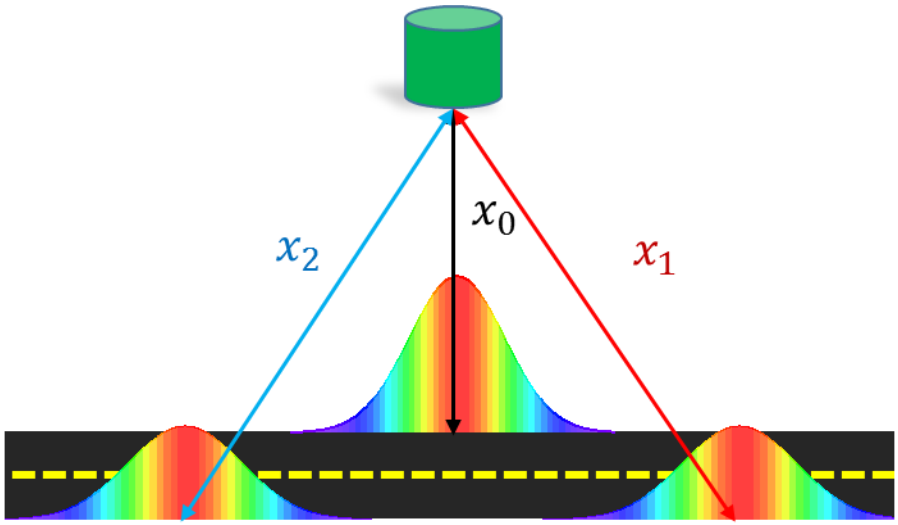
Thus, when integrating (i.e. y goes from $-\infty$ to ∞) and the transect is perpendicular to wind, the integral method becomes independent of the choice of σ_y:

$$\int_{-\infty}^{\infty} c(x,y,z)dy = \frac{Q}{\sqrt{2\pi}\,\sigma_z u}\left[e^{\frac{-(z-h)^2}{2\sigma_z^2}} + e^{\frac{-(z+h)^2}{2\sigma_z^2}}\right].$$

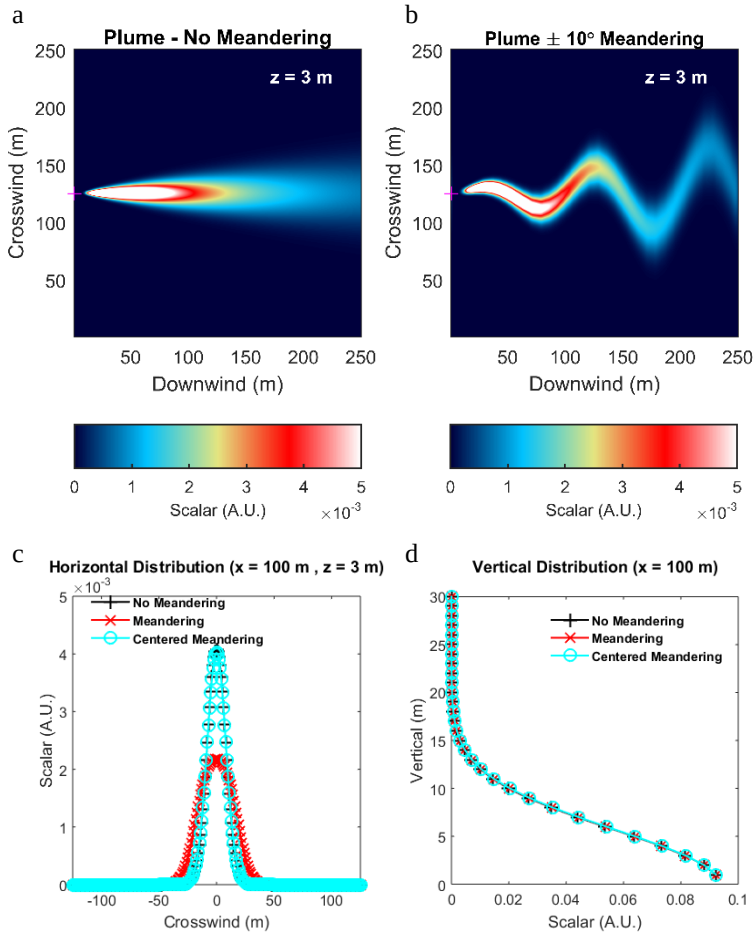
Physically, the influence of meandering as it pertains to the accuracy of σ_y becomes negligible as long as the transect captures the entire plume. Of course, this is an ideal case; transects are often not perpendicular and occasionally may not cover the entire plume width due the physical constraints of the road. This is, however, arguably a better alternative than, a regression method where the choice of σ_y will certainly affect the results. Additionally, the integration method minimizes the effect of the random variation in the plume, which could greatly affect the regression results. This approach has also been used by Albertson et al. (2016). This is also the reason why we have chosen to separate 'meandering' at longer scales from the eddy scales that contribute to the diffusion of the plume and scales that are included in either the Gaussian or the LES. This discussion has been condensed and summarized in Sect. 2.3.

2) We have now discussed the appropriateness of integration in the previous point. We have clarified in the text that we do indeed center the Gaussian to the observed concentration peak for each transect on p. 8 lines 17-20.
"Second, Eq. 1 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 the Gaussian model for each transect."

This is done for three reasons: (1) the majority (~90%) of the data processed consisted of 1-3 transects which may not be collected under conditions that match the mean and (2) the true mean wind direction is not known at most sites. For this limited amount of transects, picking an average wind direction is arguably inappropriate. Consider a case where two plumes are observed that deviate from the 'average' wind direction (shown below). In such a case the average Gaussian plume would be calculated with a distance (x_0) that is smaller than the observations (x_1 and x_2). This would underestimate the emissions.



Another way to explore this concept is to consider the average plume of all the individual Gaussian plumes and apply it to the data. An example of this was actually originally shown in Sect. 5.3 (which has now been removed due to apparent redundant discussion to newly added Sect. 4.2). Under ideal conditions, when an additional 'meander' is added to a Gaussian, the average of realizations with different downwind distances that have been re-centered will reproduce the profile of the Gaussian for the average direction. This is shown in newly added Fig. 8:



"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."

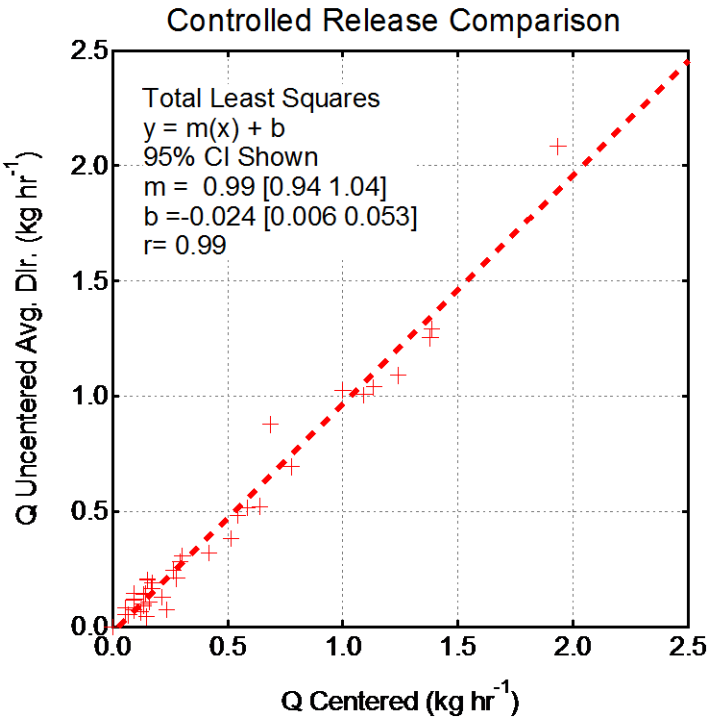
When not averaged, the profile is broader, but in any case the vertical profile is preserved. The results using the averaged Gaussian profile produced very similar results in Sect. 5.3 (1.09 vs. 1.15 kg hr^{-1}). We note here that the Referee commented on Sect. 5.3 in **point 66** saying "**As best I understand this work, the order of averaging that is varied is certain to cause no significant**

change since the relationship between concentration and source strength is linear." This approach actually *is* mathematically identical to using the average profile with all individual results (thus not centering the Gaussian on the observations), but *is not* mathematically identical to averaging the results of averaging each individual transect ratio due to the order of operations:

$$\overline{\left(\frac{\sum C_{Observation}}{\sum C_{Model}} \times Q_{ref}\right)} \neq \left(\frac{\sum \overline{C_{Observation}}}{\sum \overline{C_{Model}}}\right) \times Q_{ref}$$

We point this out to highlight that the 'best' approach is not obvious and the differences between the approaches would theoretically appear to be small. One of the goals of this work is to understand how this application of the inverse Gaussian method will affect our results with the understanding that a single or a few transects are not representative of the average conditions of the atmosphere. We expect that the error due to these differences in approaches will be encompassed in our uncertainty analysis where we examine how variable the transect-to-transect results actually are. We also concede that there are times when the individual transect centering is not the best approach, such as when information about actual source location or atmospheric conditions is a greater priority to extract from the observations.

The only method to independently verify which approach is best is to compare to a known emission rate. To that end, we have compared the results for three approaches, using the Gaussian centered to the observations, the Gaussian using the averaged profile from all centered profiles during a release period, and a single Gaussian aligned to the average wind direction during a release period. The results, shown in newly added Fig. S3, for each scenario are virtually identical. A total least squared regression for the controlled release data is shown below:



Equally important in the context of this work is the spread of the data. The results using an average profile, or profile in the average wind direction, are slightly more scattered; the centered approach gives estimates with a standard deviation 0-20% lower than the other approaches. Again, most of our data has only a few transects so this is an important consideration. The results when allowing the Gaussian to shift produce lower standard deviations, meaning any individual transect is slightly less likely to over or underestimate the mean obtained by repeat measurements.

We also not that the centering approach is not without precedent. Yacovitch et al. (2015) use the instantaneous peak location in their iterative forward modeling to estimate emissions from unknown sources and specifically state they use the bearing between the source and peak in their emissions estimates, with the reasoning the source location is more certain than the wind direction. Rella et al. (2015) effectively construct an instantaneous plume using a hybrid-Gaussian model where they measure the lowest ~50% of the profile and model the remaining portion using a Gaussian model that best fits their instantaneous measurements. Additionally, Foster-Wittig et al. (2015) presented a method that build upon the EPA OTM33A stationary method for emission estimates. In the EPA method, concentration is plotted against wind direction to produce a Gaussian shape that is then used to calculate an emission. However, this is based upon the assumptions that the plume will be in the center line and observations spread symmetrically about the wind direction. Foster-Wittig et al. (2015) presented a method to plot the data correcting for the fact that this is not always the cased which they termed a 'conditional reconstruction' of the Gaussian plume. This is not entirely analogous, but it underscores that the baseline assumption that an average wind direction can be used indiscriminately is too simplistic for real world applications and some thought must go into correcting for observations that do not match the ideal conditions. They also noted that the differences between approaches were small, errors by other assumptions of the methodology were more important to total uncertainty and repeat observations were the best way to reduce error. This discussion has been added to newly created Sect. 4.2 which includes new figures 8 and S3.

3. The final consideration here is the effect of centering the observations to the LES results, as opposed to the case of centering the Gaussian to the observations previously discussed. Because the observations themselves are integrated the same way whether they are centered or not, this is mathematically identical to not centering the results. Thus this is not expected to affect the comparison. The plume centering that we have chosen to apply to the observations is 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 meandering impact of the large scales. 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. As the LES plume is not aligned to the observations, this is another potential way to compare how the results vary when an average wind direction is used. We see generally that these results compare well, within a factor of two. Given the large uncertainties that have been reported in this method, this is considered a reasonable result. In the application of our work, the lognormal range of emissions means that even if the results

differ by a factor of two a ‘large’ emission will still be identifiable from a ‘normal’ emission. This discussion has been added to newly created Sect. 4.2.

References:

Fritz et al., 2005: https://doi.org/10.13031/2013.18501

Albertson et al., 2016: https://doi.org/10.1021/acs.est.5b05059

Yacovitch et al., 2015: https://doi.org/10.1021/es506352j

Rella et al. 2015: https://doi.org/10.1021/acs.est.5b00099

Foster-Wittig et al. 2015: http://dx.doi.org/10.1016/j.atmosenv.2015.05.042

Agee and Gluhovsky, 1999a: https://doi.org/10.1175/1520-0469(1999)056<0599:LMSTDG>2.0.CO;2

Agee and Gluhovsky, 1999b: https://doi.org/10.1175/1520-0469(1999)056<2948:FAOLES>2.0.CO;2

Horst et al., 2004: https://doi.org/10.1175/1520-0469(2004)061<1566:HFOTOS>2.0.CO;2

Summary of changes to manuscript in response to this question: Additional explanation in Sect. 2.3 (p. 7). Newly added Sect. 4.2 (p. 14) with newly added Fig. 8 and S3. Incorporation of previous Sect. 5.3 discussion into Sect. 4.2 and removal of redundant Sect. 5.3.

3. Given the limited spatial domain, the LES used in this study has no large-scale turbulence, thus its value for comparison to plume observations is uncertain. Its use for an observational system simulation experiment is acceptable, given the caveats that it only simulated neutral conditions and has no larger-scale turbulence. The authors do explain the limitations of neutral stability, but do not acknowledge that the simulation will by necessity truncate the spectrum of atmospheric turbulence that will contribute to issues like meandering of the downwind plume.

While it is true that the vertical dimension chosen will impose a limit on the large scale of eddies able to be resolved, we assume that these eddies do not contribute to the vertical diffusion of the plume. According to Townsend's theory of attached eddies (Townsend 1961), which has been confirmed experimentally by many studies (Perry and Li 1990, Nickels et sl. 2007, Woodcock and Marusic 2015 and others), the eddies that will contribute significantly to chaotic turbulent diffusion are on the order of the height of the point source above the surface. Since plume emission height and and observation height are both <<100 m, we have sufficient vertical dimension (100 m) to generate all eddies that will contribute to vertical diffusion. The larger eddies will contribute to large-scale horizontal meandering, which is why we have chosen to align the observations to the plume. Again, the larger length scales are expected to shift the plume center, but not contribute to the diffusion or spreading of the plume at a given cross-section. The discussion of the LES setup in section 2.4 has been expanded to include this discussion on p. 9 line 29- p. 10 line 8.

"It is important to note that all sites were set up to ensure that the vertical 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 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."

4. A major element of the uncertainty assessment relies on the controlled release experiments. This requires characterization of the uncertainty in the rate of release. No assessment of the uncertainty in the release rates are given. This must be addressed.

The uncertainty of the flow controller used has been added to the text in Sect. 4.3. on p. 16 line 11-12.

"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)."

The flow controllers were calibrated with a Gilibrator soap bubble meter and no differences were observed.

5. The measurement protocol chosen focused on morning and evening conditions to search for neutral atmospheric stability. Sounding on calm mornings, however, are far neutral, and morning and evening conditions are characterized in general by rapidly changing atmospheric stability. Plume dispersion modeling assumes steady atmospheric stability and mixing conditions. What is done to screen the observations to avoid rapid transitions in stability that clearly violate the conditions for the Gaussian plume model?

Additional discussion has been added to Sect. 2.3 to clarify the limitations of generalizing this analysis to rapidly changing conditions on p. 8 lines 8-13.

"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."

We have addressed the expected error of miscategorizing the stability in Sect. 5.1. on p. 19 lines 20-31.

"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 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 +/-1 stability class is relatively consistent at farther downwind distance, averaging a change of 40% at 200 m downwind."

Detailed comments.

1. Page 1, Line 14-15. " . . .a hierarchical sampling with increasing complexity was implemented" is, I believe, what you mean to say. Though I'm not sure what "with increasing complexity" means. It is hierarchical, thus complex.

We have attempted to clarify this and several related points of confusion by remaking Fig.6 to be more informative.

2. Page 1, transects, lines 15-20. How are multiple transects separated in time? Across multiple days? Or all on the same day / within the same hour? This is an important methodological distinction.

The text in the abstract on p. 1 lines 15-17 has been modified to *"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 in 15-60 min. "

In addition, Table 2 describing the number of samples and their timescales has been added and this information has been added to Fig. 6.

3. Page 1, Line 20. "in most cases with average differences" Be more precise. Is 25% the average difference or not?

This line has been clarified to indicate this in the average difference on p.1 line 21-22.

"The LES and Gaussian derived emission rates agreed within a factor of 2 in all except one case; the average difference was 25%."

4. Page 1, Line 24-25. "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." This is not very informative. What is relationship between number of transects and precision of determination of the emission rate?

We have added additional text on p. 2 lines 2-4 to explain the uncertainty range.

"*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."*

5. Page 1, Line 25-26. "In addition, other sources of uncertainty including source location, wind speed and stability were analyzed." This is too vague to be informative.

The text has been changed on p. 1 lines 28-29.

"Sources of uncertainty from source location, wind speed, background concentration and atmospheric stability were also analyzed." However, we have tried to streamline the abstract for the most important points. Thus, we believe that this description is sufficient for the abstract.

6. Page 1, Line 26. "atmospheric variability" is too vague. Do you mean to say sampling error caused by atmospheric turbulence? If so, isn't that a function of the sampling strategy? Is atmospheric sampling still the dominant source of uncertainty with 20 downwind transects?

We have added additional text on p. 1 lines 28-31.

"*The largest contribution to the total uncertainty was from atmospheric variability; 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."*

7. Page 1, Lines 28-30. What is "this condition?" What is "this metric?"

Both instances have been replaced with "atmospheric variability" on p. 2 line 1 and p. 2 line 4.

8. Page 3, line 1. "lag behind the standard." In what respect?

This section has been clarified on p. 3 lines 25-27.

"However, as seen in Table 1, current results using ground based and mobile measurements seem to underperform relative to other measurement techniques in this field in providing low uncertainty emission estimates."

9. Page 3, equation 1. Please incorporate the equation into the English, and please define C and the origin.

C is now defined on p. 4 line 6. *". . . scalar concentrations (C). . ."*

We have now defined x and the origin on p. 4 lines 12-13.

"Here, the source is defined as the origin with x the downwind distance, y the crosswind distance and z the measurement height above ground."

10. Page 3, line 15. "Gifford's"

The apostrophe has been added on p. 4 line 16.

11. Page 3, line 16-17. It is not necessary to describe the stability classes. The model does not predict a PDF of the concentration. It predicts the average concentration. These are very different.

The description of the Gaussian plume has been clarified on p. 4 lines 17-18.

"The model describes the distribution of a scalar concentration (C) downwind of a source, meaning it describes average plume locations and concentrations."

We believe it is suitable to leave the description of stability in for less experienced readers to have a full understanding of the Gaussian parameterization.

12. The uncertainties in Table 1 are not helpful as currently presented. This Table must be revised prior to publication. I understand that they are all reported by the individual authors, but this manuscript does present them for the purpose of comparison. Many of the techniques and papers have radically different spatial and temporal domains and sampling density, and the uncertainties quoted are thus not at all comparable. How, for example, can aircraft estimates of the mass balance for an entire basin be compared directly to the uncertainty of a single chamber sample, or a single Gaussian dispersion estimate, without an understanding of what this uncertainty bound is trying to represent? This table could be enhanced to include some information about the domain and sampling density associated with these studies. This would be challenging, but much more interpretable.

We have added additional clarifying information in the text pertaining to the different scales that some of these techniques could be used at to encourage readers to investigate techniques more thoroughly if they are interested in a specific application. Text has been added to p. 3 line 3-9.

"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 is intended to illustrate the motivation for this work and not as a comprehensive review of all these methods. While we agree that caution is needed in applying and comparing these methods, many methods have been used at different scales even in the same category making it difficult to add this information.

13. Page 4, lines 3-12. Are these uncertainties for single emissions estimates from a site using an individual "transect" of data? Please clarify. See my concerns above. Without specifying more about the temporal and spatial domains, these uncertainties are difficult to compare or interpret.

The text has been clarified on p. 5 lines 12-14.

"These methods report uncertainty analyses that 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."

Authors describe their error estimate in terms of 'sites' which does not distinguish estimates for sites with different density of data collected. We will assume this is meant to apply to sites with single or multiple transects.

14. Page 4, line 19. What is "an averaging plume for unconstrained instantaneous measurements"? Please clarify. I study boundary layer meteorology but do not understand what you are trying to say.

The text has been clarified on p. 5 line 28-29.

"Such situations would include using the average Gaussian plume model with instantaneous measurements without uncertainty or sensitivity analysis or applications over complex topography."

15. Page 4, line 25. I don't understand, "this method." Are the authors proposed a new standardized methodology for quantifying emissions? Or is this a study of the uncertainties in plume measurement / source quantification methods?

The text has been clarified on p. 6 lines 4-7. *"The basic architecture of our approach to quantify errors associated with mobile Gaussian methods uses (1) advanced modeling of a preselected sample site to enable investigation of optimum sampling strategies, (2) application of strategies to the sample collection process and (3) evaluation of additional sources of uncertainty and bias using advanced modeling and a controlled release."*

While we necessarily have to explain our own methodology and uncertainty analysis to support our own studies, we have tried to extend the usefulness of such an endeavor to the broader community by making general recommendations for the best practices of this method. We have added additional descriptions and reworked Fig. 6, combining it with the previous Fig. 7, to clarify out workflow.

16. Page 5, line 5. The methane span gas concentration is quite low for the measurement of plumes at close range. Are there any tests of the linearity of the instrument at higher concentrations that I assume are encountered driving downwind of strong methane sources?

The LI-COR 7700 has a stated measurement range accurate up to 40 ppm. The average enhancement that we saw was only ~1.25 ppm above background. Given the stable performance of the LICOR, this calibration standard should not adversely affect the LI-COR in the ambient range of our observations (McDermitt et al. 2010). The text has been modified on p. 6 lines 18-20.

"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. 2010). In addition, enhancements observed were usually less than a few ppm above the ambient concentration."

17. Page 5, lines 19-20. "site were screened": Does this mean that sites with trees were eliminated from the sample? The text later states that it is hard to measure a plume at a distance greater than 300m if trees are present. Can you please explain the site selection algorithm more carefully?

Sites with trees were not ideal and generally not added to our database of possible measurement sites. However, practically this was difficult to achieve. Sites with full tree cover were avoided during this screening process, but sites with a few trees were not as this would have been too strict a threshold. The text has been clarified to discuss site screening on p. 7 lines 4-6.

"Sites were screened to remove those that were far (>300 m from public road), had obstructions (buildings and full tree lines) and large elevation difference (>50 m)."

18. Page 6, line 7-9. Does the NOAA web site provide observations, or numerical model reanalyses? Please clarify. How does this NOAA web site account for the impact of local roughness on atmospheric stability? The hilly, forested, heterogeneous landscape of Pennsylvania will have a significant impact on local dispersion.

NOAA produced meteorological data from a climate analysis model with data assimilation. The text has been clarifies on p. 7 line 31-p. 8 line 8.

"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. 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 at 40 km resolution and constant pressure coordinates. The data were interpolated to 1 hour resolution for use in the model."

The NOAA uses a model with an observation network as inputs. More information can be found on their specific website given in the text above. Surface data such as vegetation type and soil are included and they do provide data on surface validation. However, the 40 km scale likely diminishes some aspects of the surface inputs. Note that due to the extensive nature of this study we chose this scale as it was applicable for a larger area of our sample and minimized the number of different inputs for a given day.

19, Page 6, equation 2. As with equation 1, please make equation 2 part of the sentence.

The text has been changed accordingly.

20. Page 6, line 20. This description is insufficient. The Gaussian plume model describes only the concentrations caused by the local source. The measurements do not. A concentration background must be defined and subtracted from the observed time series. This is a critical and non-trivial step. How was this done? I am also concerned that this method does not take into account the correlation between the modeled and observed concentration enhancements. Systematic errors in the dispersion would not be identified by simply matching the integrated downwind enhancement.

We believe we have addressed background removal in Main Point 1.c. The use of LES to understand the pitfalls of Gaussian dispersion is a major focus of this work. More generally, the discussion of the advantages of integration are addressed in the response to Main Point 2.

21. Figure 2. This figure is concerning. First, the figure shows concentrations that differ by a factor of 1000, and no explanation is offered. Second, the background that must have been removed from the observations is not discussed. Third, there appears to be a systematic difference in the width of the observed vs. modeled plume. This implies a systematic error in the dispersion coefficient, which will lead to a systematic error in the emissions estimate. How is this addressed? Finally, given that instrument characteristics are expressed in units of ppb, it would be useful to also present downwind concentrations in these units. It would be easy to include both mass density and mole fraction as two different y axis labels.

Figure 2 has been clarified to explain the model is simulated with a 1 kg/s source in the figure caption and the ppm axis has been added. In addition, the text on p. 8 lines 19-23.

"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 the Gaussian model for each transect. A comparison of observations and model output 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."

We believe we have addressed the question of the background in Main Point 1.c and have also included additional clarification in this section on p. 8 lines 15-17 and Eq. 2.

"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."

We have not intended to neglect the importance of plume dispersion, but have chosen to not delve into it at this point. We have discussed horizontal and vertical plume dispersion differences between the Gaussian and LES later in Sect. 4.4. We reiterate that the cross-wind integration removes any influence of the lateral dispersion of the plume on the inferred emission rate.

22. Page 7, line 11. How is steady-state evaluated?

The text has been clarified and the relevant figures provided in the supplemental have been referenced in the text on p. 10 line 9-11.

"Sites were simulated for at least 30 min to allow the simulated turbulence to reach a statistically stationary state, where average and standard deviations of wind and scalars approach a constant value as shown in Fig. S2."

23. Page 7, line 11. What does it mean to "rescale" the LES output to the measured momentum flux? Isn't the LES set up to simulate these sites? Why is rescaling necessary?

The numerical simulations of LES are generally non-dimensionalized for simplicity. The text has been clarified on p. 10 lines 11-13.

"The equations solved in LES are non-dimensionalized using the friction velocity (u_*). The advantage is that results from LES apply more broadly to any problem when the non-dimensional quantities, e.g. as $u_{nondimensional}=u/u_*$, are considered."

24. Section 3.1 The LESs are truncated in the vertical and thus cannot include largescale turbulence that will influence the sampling statistics. This will be particularly important for unstable conditions. How is this dealt with in designing the sampling strategy?

We have addressed this in Main Point 3. Additional discussion and references for have been added in Sect. 2.4 to encompass the theory and work that has proven that the important scales of eddies are ~10x the perturbation allowing a smaller boundary layer to be suitable for this analysis. LES was not used to investigate stability as indicated in Table 3, which has been added to the main text.

25. Section 3.1. The LESs say nothing about sensible heat fluxes. Turbulence statistics change dramatically as a function of stability conditions. What stability conditions are simulated?

This has been clarified in Table 3. Only neutral conditions are simulated.

26. Page 7, line 25. Short intervals will sample the same structures in the turbulence. This is well known. Some of the existing literature on sampling statistics in atmospheric turbulence should be cited here.

This is precisely why multiple transects are needed, and why they need to be spaced by some minimal time to avoid resampling the eddies. Additional citations have been added on p. 10 line 29-p. 11 line 2.

". . . however, the time interval may also affect results as measurements with short spacing may resample the same coherent plume and thus the same plume realization (Metzger et al., 2007, Shah and Bou-Zeid, 2014)."

27. Page 7, line 26. What does "1-N" mean? What is N? I read the description, but I do not understand what this means. What defines the number of samples available in a numerical simulation?

The text has been clarified to point out the LES time resolution (1 Hz) imposed for the sake of minimizing file size and remove the ambiguity of the value 'N' on p. 11 lines 3-7.

"Using the LES output, which is saved with 1 Hz resolution to match our instrument sampling frequency, sample transects were picked from the full time series, varying the number of 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."

28. Page 8, lines 8-9. What is the difference between turbulence and plume meandering? Isn't plume meandering caused by atmospheric turbulence?

See comments in Main Point 2. We have added Sect. 4.2 on p. 13 to specifically address this question and clarify that while meandering has been used to describe plume dispersion more generally, we are specifically using it to refer to larger scale motions. A fundamental difference is that meandering does not increase the cross sectional area of the instantaneous plume since it is effected by scales larger that the plume diameter (scales that simply move and instantaneous plume around). Eddies with sizes similar to the plume diameter diffuse an instantaneous plume.

29. Page 8, lines 10-12. What does it mean to "reflect the actual variability in the atmosphere"? This is not meaningful. Please define a quantitative tradeoff between transects and flux retrieval.

The text has been clarified on p. 12 line 2-7.

"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."

The quantitative tradeoff is reflected in Fig. 5, which shows reduced variability in the returned emissions estimates with increasing numbers of passes.

30. Figure 3 is not very helpful. Please include vegetation and topography.

Figure 3 has been changed to include satellite imagery.

31. Figure 4. Images of instantaneous vs. averaged plume structures do not need to be published. This is well known. The manuscript has no references to several decades of simulations of plume dispersion. This is a serious hole in the scholarship of this manuscript. Look up, for example, the publications of Jeffrey Weil, circa 1990.

Additional references have been added to Sect. 3.1 on p. 10 lines 21-23.

"LES 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 (Nieuwstadt and de Valk, 1987, Weil 1990. Wyngaard and Weil, 1991, Mason 1992, Weil et al. 2004)."

We feel the figure is instructive for those with limited exposure to LES, thus we have opted to keep the figure in. This manuscript is not intended to be a literature review thus we have limited our discussion to the most relevant papers.

32. Figure 5. Issues.

1. There is a lot of blank space.

2. The percent difference figures show results from a distribution. What are the elements of the distribution? Is each "pseudo-transect" an element of the distribution? Please define this clearly in the caption.

3. A standard deviation is a description of a population. Are there multiple populations whose statistics are being compared? Please clarify.

4. What does "random seconds" mean? Why is this useful?

5. There is little additional information given in the 30/60/120 second sample repeat figures, and the differences are difficult to see. It would be more effective to show one of these, then have one additional figure showing the differences caused by changing the sampling interval.

Figure 5 has been condensed for clarity and the original figures have been moved to the supplemental as Figure S4. The text has been clarified on p. 11 lines 9-16.

"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."

33. Page 8, lines 18-19. The original plan is not relevant at this point. Present the number of samples obtained.

The text has been adjusted on p. 12 lines 13-16 to:

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

The actual sample numbers were already available on p. 12 lines 17-18.

34. Page 8, line 24. What is the "average maximum percent difference"? Maximum among what? Average across what? And why is this relevant? Please also explain the populations used to define the standard deviations.

The text has been clarified on p. 12 line 21-26.

"For the population of standard sampling sites with multiple passes, the average rsd of 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."

35. Page 8, last sentence. What is the population being discussed here? Why are these numbers important?

The importance of the LES derived values, which we are comparing to, has been clarified in Sect. 3.1. For each combination of transect number and time spacing, 100 random samples were picked and the mean and standard deviation of the emission rate was calculated. 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 time simulation and are not likely to resample similar plume structures. The text has been clarified with additional discussion on p. 12 line 21-29.

"For the population of standard sampling sites with multiple passes, the average rsd of 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."

36. Figure 7 does not explain how the number of sources will be determined. I also disagree with the far right-hand bar. A more complex simulation does not ensure less uncertainty, and "model uncertainty" is not defined. If there is a precise definition of uncertainty that can be shown to be reduced with the LES, please explain and define this.

Figure 7 has been combined with Figure 6 and redone for more clarity in what is now Figure 6. The combined figure shows the full workflow including that sites with more measurements were also used with the higher complexity model. The uncertainty column is meant to be representational of the expected reduced uncertainty due to more available measured input information, not necessarily the model complexity. Clarifying discussion has been added to p. 13 lines 15-16.

"*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."*

The source locations are discussed in the text as it would be difficult to succinctly portray in a figure. See comments to point 41.

37. Page 9, line 17. What is a "warm up time"? (I see here that the determination of steady state is explained.)

The text has been clarified on p. 13 line 23.

"LES time series spanning ~30 minutes were averaged to produce a pseudo Gaussian distribution excluding a ~5 min initialization period (time until stationary state is achieved)."

38. Page 9, line 21-24, and figure 8. This discussion and figure displays a fundamental problem with the LES setup, and the authors' interpretation of these data. The "real world" meandering of the wind that is being described is turbulence in the atmospheric boundary layer. A Gaussian plume model describes the average state downwind of a point source, which would be properly represented by the average of all of the transects, without "aligning" them to match the peak concentration on each transect. Aligning these is the equivalent of performing some sort of high-pass filter.

See comments in the response to Main Point 2.

39. Similarly, I am now concerned about the processing of all of the transect data. In Figure 2a, for example, have all of the transects been arranged so that the x direction changes from transect to transect? If so, the same odd filtering of turbulence has been applied. This is not appropriate for comparison to a Gaussian plume model. I expect this could explain the difference in plume widths that is evident in Figure 2 (note the modeled plume is wider that the observed plumes that may have been filtered to remove large-scale turbulence).

See comments in the response to Main Point 2.

40. Please explain the theoretical basis for equation (3)?

Equations 3 and 4, which dealt with the proper dimensionalization of the LES non-dimensional output to calculate and emission rate have been combined and explained on p. 13 line 25 – p. 14 line 8 to improve clarity.

"Because all LES runs on non-dimensionalized, the LES output needs to be scales to represent the actual field conditions for a given observation. The LES 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:

$$\frac{C_{LES}}{\dot{M}_{LES}} \times u_{LES} = \frac{C^D_{LES}}{\dot{M}^D_{LES}} \times u^D_{LES} \qquad (3)$$

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

$$C^D_{LES} = \frac{\dot{M}^D_{LES}}{\dot{M}_{LES}} \times \frac{u_{LES}}{u^D_{LES}} \times C_{LES} \qquad (4)$$

Here, $\dot{M}^D_{LES}$ is the normalized mass (1 kg s^{-1}) and $\dot{M}_{LES}$ is the actual non-dimensional emission rate introduced into the LES system. Normalizing by $\dot{M}_{LES}$ is necessary to match the reference emission rate (Q_{ref}) also used in the Gaussian retrievals. The non-dimensional wind speed (u_{LES}) is divided by the dimensionalized (u^D_{LES}) wind speed which is set equal to the on-site tower measurements."

41. Section 4.1. How do you distinguish among single vs. multiple sources? This is not explained.

The text has been clarified to explain how sources were identified on p. 13 lines 16-19.

"*As on-site access was not available, all large structures were treated as sources. These were visually identified and exact coordinates were confirmed in Google Earth. The identified sources were always gas processing units or storage tanks."*

42. Page 10, line 2. The two Gaussian retrievals "compare"? What does this mean?

The text has been clarified on p. 16 lines 15-19.

"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; NOAA overestimated the winds during the latter two release rates."

43. Figure 9. Once again, please explain the populations of data that go into these box and whisker plots. Is each point a transect? If so, how many transects make up each population?

The populations of Fig. 9 are now defined on p. 16 lines 14-15.

"Boxplots of the populations of emission retrievals from three scenarios are shown in Fig. 9 . . ."

We have moved supplementary Table S1, which contains site specific details, to the main text as Table 3 and referenced it in the text on p. 16 lines 12-13.

"Site set-up was discussed in Sect. 2.2 and specific site details are available in Table 3."

Specific numbers of transects for each release have been added to Table 4.

44. Page 10, line 13-14. I do not agree that the results show no apparent bias. Figure 9a has winds that are known to be too high, according to the text. Figure 9b shows a systematic underestimate of emissions. Figure 9c shows varied results.

The text has been clarified 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."

As shown in Table 4, panel b does not consistently underestimate the emissions. The issues concerning panel c are discussed in the text on p. 16 line 30-p. 17 line 3.

"While LES can readily account for unstable or stable conditions, this would come at an increased computational cost as multiple simulations would be needed. The results here show that this 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."

45. Figure 10. What is the level of uncertainty in the release rate? If this release is being used to evaluate the methodology, then the release rate must be carefully calibrated. How has this been done?

The release rate uncertainty has been added to the text on p. 16 line 11-12 and Table 4.

"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)."

The flow controllers used were factory calibrated and tested in the lab before use.

46. Page 10, line 22-23. Only one outlier with zero emission is shown in Figure 10.

In release 1, 0 is not considered an outlier given the spread of the data. The text has been updated on p. 17 lines 11-12.

"During release 1 and 3, even though a constant source is being emitted, a few (1-3) transects showed no observed plume whatsoever."

47. Table 2. Please report the observed and simulated winds, and the observed and simulated stability conditions. Please include more descriptive information in the table caption. In particular, please note how many transects were collected at each site, the time span of the transects, and the source of the uncertainty values given in the table. If these are standard deviations based on emissions derived from individual transects, and the purpose is to compare mean values, the authors should report the standard deviation of the mean, not the standard deviation.

We have moved supplementary Table S1 to the main text as Table 3 (which contains time intervals, transect # etc.) and referenced it in the text to mitigate confusion. What is now Table 4 has been updated with the standard deviation of the mean as well as transect numbers.

48. Table 3. Please include more descriptive information in the table caption. In particular, please note how many transects were collected at each site, the time span of the transects, and the source of the uncertainty values given in the table. If these are standard deviations based on emissions derived from individual transects, and the purpose is to compare mean values, the authors should report the standard deviation of the mean, not the standard deviation. Also, please report winds and stability. And were the winds observed, or taken from NOAA reanalyses?

What is now Table 5 (formerly Table 3) has been updated with standard deviation of the mean and the number of transects. In addition Table 3 has now been moved to the main text with site specific information. Table 2 has also been added to more generally explain data available for each level of data.

49. Page 10, lines 27-28. See my earlier concerns about centering of the plumes on each transect. I believe this is an erroneous interpretation of plume dispersion.

See comments in Main Point 2.

50. Page 11, lines 7-8. The overall differences are small? The flux estimates in Table 3 differ by as much as a factor of two. This seems large to me.

The text has been clarified on p.18 lines 1-2.

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

We would also clarify that while this may seem large, emissions may span many orders of magnitude so this level of comparison is actually quite good for this method.

51. Figure 11. This figure points to many questions and problems.

1. **What are the equations for the vertical and horizontal flux profiles in the Gaussian plume model?**
   An appendix with the pertinent equations has been added on p. 25-26.
2. **What is the position in the domain where the profiles (figure d-g) are computed?**
   The figure caption has been updated to indicate that these are computed for the road plane (as shown in Fig. 3 with distances specified in Table 3).
3. **What is the distance downwind for everything shown in this figure?**
   The figure caption has been updated to indicate that these are computed for the road plane (as shown in Fig. 3 with distances specified in Table 3).
4. **How are the multiple sources chosen? As best I can tell this is never described anywhere in the manuscript.**
   See comments in point 1b.
5. **Does the LES flux include subgrid and resolved fluxes? Please explain, and delineate these two sources so that it is clear what fluxes are explicitly resolved.**
   Only resolved fluxes are shown in these figures. Subgrid scale (SGS) contributions are important for higher order statistics, but when dealing with the mean concentrations they are not directly relevant. That is, for the mean concentrations used to infer emissions, the SGS contribution is theoretically zero. For this figure, the horizontal flux is a mean advective flux and there is no SGS contribution, but for the vertical flus an SGS contribution exists and is modeled in the LES, but only the resolved flux is plotted here. The text has been clarified in the Fig. 11 caption.
   "The vertical LES flux corresponds to the resolved fluxes only."
6. **What does z=3 mean? 3 what? The same comments and consternation hold for the similar figures in the supplementary materials.**
   Z is the vertical altitude as consistent with panels d and e. The title has been clarified to add the 'm' unit.

52. Page 11, lines 12-14, sentence starting with "regardless." I do not understand what you are trying to say.

We have changed the text on p. 18 lines 5-12 to:

"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, 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 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."

53. Table 4. What is the distance from source of the measurements? Are "emissions" those estimated from transect measurements? How many transects? What are the meteorological conditions for these measurements?

The supplemental table with this information (number of transects, distance, meteorology etc.) has been moved to the main text as Table 3. Table 4 has been removed as it was intended to be an example and not a systematic analysis as described later in the text.

54. The text in the first paragraph of section 5.1 does not appear to match the results in Table 4. The text notes that unless the change in source location in the x direction is similar to the distance of the measurement downwind, that the impact of the estimated emission is small. Table 4 shows two examples, one of which has a 150% change in source strength estimate, and another a 7.5% change in source strength estimate, and as best I can tell (Table S1?) the measurements are both from about 150m downwind of the source, with the site with a larger % change having measurements farther downwind. Based on the results presented by the authors, as best I can interpret them, I disagree with their conclusions.

Table 4 is a simple exploration of 2 samples sites where as Figures S11 and S12 actually systematically look at the effect of a specified uncertainty with changing downwind distance and height for different downwind measurement distance. Table 4 has been removed to reduce confusion and streamline the discussion.

55. Figure S10 (b) is uninterpretable. What is the "Ratio between the sum in y and distance x of the scenarios"? Is this truly a ratio of distances that is being plotted? Please clarify.

What is now Figure S12 has been redone to present data as a percent reduction from the original model to improve clarity. The percent change in the model is equivalent to the change in the calculated emission rate for these scenarios.

56. Figure S11 (b). See comment above regarding Figure S10.

What is now Figure S13 has been redone to present data as a percent reduction from the original model to improve clarity. The percent change in the model is equivalent to the change in the calculated emission rate for these scenarios.

57. Page 11. "NOAA wind speeds differed from the tower data on average by 50%." This is uninterpretable. Please rewrite this to be meaningful.

The text has been clarified on p.19 lines 6-7.

"In the context of this analysis, we compared the NOAA wind to the mean on-site tower measurements of winds at 16 tower sites. NOAA wind speeds reported higher and lower values than the mean tower winds; the absolute difference was 50% on average."

58. Figure S12. Please define this ratio, as with Figures S10 and S11.

Figure S14 has been redone to present data as a percent reduction from the original model to improve clarity. The percent change in the model is equivalent to the change in the calculated emission rate for these scenarios.

59. Page 12, lines 7-8, "The magnitude of the difference between consecutive stability classes is relatively consistent, averaging 40%." Figure S12 does not suggest simply defined average, as the signs and magnitudes of the differences change as a function of distance downwind. Please explain what this 40% value means.

Figure S14 has been replotted to clarify this point. The text has been clarified on p. 19 lines 29-31.

"The magnitude of the difference between +/-1 stability class is relatively consistent at farther downwind distance, averaging a change of 40% at 200 m downwind."

60. Page 12, lines 12. Please define "absolute terrain slope."

The text has been updated on p. 20 lines 2-3.

"The geometric mean of the absolute terrain slope, defined as the absolute value of terrain rise over the distance between sources and observation . . ."

61. Page 12, lines 24-28. The methods for quantifying the uncertainties should be explained carefully in the methods. This rapid-fire, qualitative overview of the methods behind Table 5, arguably the most important result of the manuscript, is insufficient.

The details of the uncertainty work are described throughout the manuscript. Sect. 5.1 has been expanded to clarify many points including the separate uncertainties for source location and height. Discussion of the Monte Carlo approach are also expanded in this section.

62. Page 12, lines 30-34. See previous concerns about the lack of larger scale turbulent motions in the LES used for this study. I agree in general that the LES used here should provide a lower estimate of turbulent sampling error, but it also will not contain the full spectrum of atmospheric turbulence and true turbulent sampling error. This should be explained, and references to the rich literature of the full spectrum of atmospheric turbulence and the limitations of LES that is limited to the atmospheric surface layer should be included in the manuscript.

See comments in Main Point 3. It is true that LES will not resolve the entire spectrum of atmospheric motions, but as we explain, the motions important for turbulent diffusion are resolved. In addition we have added the following text and references to p. 20 line 29-p. 21 line 3:

"*LES has a limited ability to represent the very-large scale motions (Kunkel and Marusic, 2006) or 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. "*

63. Page 13, line 6. The ranges of the input values used in the Monte Carlo simulation and their distributions must be described, or these results are meaningless.

Table 7 has been added with inputs and additional text has been added on p. 21 lines 9-20.

"To combine the remaining sources of error, a Monte Carlo 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 uncertainties may not be normally distributed and emissions are constrained to be above zero, causing a skew to the emission retrievals. A generic scenario matching average conditions experienced during the measurements was devised using a downwind distance 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 emission rates as shown in Fig. S15."

64. Table 5. See prior concerns about the lack of documentation of the methods for assessing these uncertainties. In addition, how is the "total" uncertainty assessed? Is this the result of the Monte Carlo simulation described in the text? If so, see prior concerns about documentation of this experiment. If not, please explain how this total is computed.

See comments to point 63 to address Monte Carlo documentation. We have removed the total uncertainty line and added Table 7 to display the 95% CI for the scenarios.

65. Figure 12 results vary by a factor of 1000 with no explanation.

The caption for Fig. 12 has been updated to explain the scale difference.

66. Section 5.3. As best I understand this work, the order of averaging that is varied is certain to cause no significant change since the relationship between concentration and source strength is linear. Thus this comparison is not informative or significant. If anything is nonlinear that could cause a difference, please explain.

Due to the expanded discussion in Sect 4.3 we believe Sect 5.3 is now repetitive and have opted to remove it to streamline the manuscript.

67. Page 14, lines 14-15. "Likewise, the single transect method, though fast, has many sources of uncertainty. Hence, the combination of all of these approaches described in Sect. 3 is expected to maximize sampling efficiency while minimizing uncertainty." I do not see how your results justify this statement.

By explaining the Monte Carlo approach we have shown that there is value in using any and all of these approaches (single transect, multiple transect and LES) to reduce uncertainty. We have clarified the text on p. 22 lines 27-30 to:

"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."

68. Page 14, lines 22-23. I do not yet agree with the assessment that no bias is observable. This cannot be assessed until the uncertainty bounds in Table 2 are clarified. If the mean values differ significantly, then there is observable bias.

See comments in point 44. We have also clarified the text on p. 23 lines 7-10.

"Using the LES and a controlled release, we confirm that the Gaussian model performs well 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 expected on average."

69. Page 14, line 23,"LES is therefore not required for studies where source strength calculation is the main goal." The authors have clearly avoided complex terrain where LES is most likely to be needed. This broad and general statement is not justified by the research presented in this manuscript.

The text has been amended on p. 23 lines 10-11.

"*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."*

70. Page 14, lines 24-25. "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–6.0x where x is the emission rate." This is a potentially important result, but this is offered as a passing comment. The methods behind this calculation are not presented. This is not acceptable for publication. This is an important result whose methods must be clearly explained and defended.

Table 7 has been added with inputs and additional text has been added on p. 21 lines 9-20.

"To combine the remaining sources of error, a Monte Carlo 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 uncertainties may not be normally distributed and emissions are constrained to be above zero, causing a skew to the emission retrievals. A generic scenario matching average conditions experienced during the measurements was devised using a downwind distance 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 emission rates as shown in Fig. S15."*

71. Page 15, lines 3-4. Rella et al, (2015) did not use a Gaussian plume dispersion approach, and attempted to measure the entire vertical extent of the plume. This manuscript notes that vertical dispersion is a major source of uncertainty in their results. It seems very likely that Rella et al's approach should, therefore, yield a significantly smaller uncertainty estimate than a method that relies on a Gaussian plume dispersion model to quantify vertical dispersion.

The text has been clarified to include discussion of their vertical plume information. p. 23 lines 24-27.

"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."

Though Rella et al. do have vertical information it is not enough to resolve the entire plume, thus they still rely on a Gaussian plume equation to calculate emissions.

72. Page 15, lines 5-6. "As described in Sect. 3.2, the observed atmospheric variability can range from 10-200% meaning that in-situ observations of variability and post screening out conditions with unacceptably high variability may be a viable way to reduce uncertainty." What is the quantity "atmospheric variability," that can range from 10-200%, and 10-200% of what? What criteria of "unacceptably high variability" can be used to screen out data and thus reduce uncertainty? This reasoning is either explained poorly or imprecise, and the conclusions stated are thus either uninterpretable or inaccurate. This statement is not suitable for publication.

The definition of atmospheric variability has now been clarified as discussed in Main Point 1a. The text has been clarified on p. 23 lines 28-p. 24 line 3.

"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 from repeat measurements at a single site may potentially be used as a post screening method to exclude conditions that will lead to extremely high uncertainty in single transect sites."

73. Page 15, recommendations. Point 1 cannot be satisfied in many cases. What happens when measurements are needed for a location that does not satisfy these criteria? I don't understand point 5.

We have clarified point 5 on p. 24 to:

"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 constrain atmospheric variability which is expected to be the largest source of the uncertainty estimate."

We believe we have presented sufficient evidence in defense of point 5 now with the expanded discussion of the Monte Carlo approach and how uncertainty ranges change for multiple and single transects. The Referee's comment on Point 1 is valid, that is why we make a point to clarify this is best practice and necessary to compare to our results.

74. Point 6 is not a significant recommendation resulting from this research.

In this manuscript we looked at uncertainty from stability, source location, source height, wind speed, dispersion model and number of transects. Previous studies have published work that rarely accounted for any of these. Therefore we feel it is always useful to reiterate the importance of planning for uncertainty analysis and have opted to leave this point in as it remains a valid, and forgotten, recommendation.

75. I do not understand "the strategy" proposed in Point 8. It is not clear how this collection of measurements is proposed as an integrated sampling strategy.

We have attempted to clarify this throughout the text and with Fig. 6. which shows more succinctly how the workflow is divided for data of different time intensity. We emphasize again that multiple transects are needed to quantify atmospheric variability, which is the single largest source of uncertainty. Routinely measuring the state of the random error imposed by the atmosphere will allow better uncertainty analysis for sites with fewer transects.

76. Section 8. Data should be publicly available, and not restricted to access only via correspondence with the authors.

As this dataset is expected to be reference by multiple works, a data file containing the emission rates, date/time, 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 and will provide a stable url where the data can be accessed. The URL location of the dataset is now explicitly mentioned in the supporting information. Note that the University understandably makes it very difficult to modify datasets one posted, and for this reason, we have held off on publishing the data until the review process is finished.

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

~~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 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 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 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 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, 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 (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.

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:

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

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.

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 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. 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 (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.

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. 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 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

~~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 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. 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 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 final design of the mobile lab, the instrumentation rack and the mobile tower are~~is~~ 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~~,~~

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~~.~~:

$$Q = \frac{\Sigma C_{Observation}}{\Sigma C_{Model}} \frac{\Sigma C_{Observation} - C_{Background}}{\Sigma C_{Model}} \times Q_{ref} \quad (2)$$

~~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

a short time scale, even when a few transects are averaged. It should also be noted that under ideal conditions (e.g. road perpendicular to prevailing wind direction), integrating the Gaussian equation (Eq. 1) in y creates a Gaussian profile independent of the choice of horizontal diffusion (σ_y). As the vertical diffusion (σ_z) is expected to be independent of averaging time (CCPS, 1996), this also has the advantage of minimizing the effect of different measurement timescales when 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 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 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 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 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 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 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 steadya 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 outputare non-dimensionalized using the friction velocity (u_*). The advantage is rescaledthat 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 allow that allows direct comparison to the Gaussian model estimates. Table S12 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

3.1 Model-Based Design of Sampling Strategy

As the LES outputhas 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 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

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:

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

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:

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

Here, $\dot{M}^{D}_{LES}$ 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^{D}_{LES}) 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.

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

~~$$Q^{*}_{LES} = \frac{u^{t}_{LES}}{u_{LES}} Q_{LES} \quad (4)$$~~

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

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 ± 0.01 kg hr^{-1}, 0.~~22~~216 ± 0.002 kg hr^{-1} and 0.~~09~~090 ± 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 ~~investigate~~assess 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 large~~an 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

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

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

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 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. 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 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 ~~5~~6. 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 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 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 to~~the effect of changes in the mean flow and meandering plumes ~~which~~that can contribute significantly to overall atmospheric variability (Vickers, Mahrt and Belusic, 2008, Mortarini et al., ~~2016). This is also the practical~~2016). LES has a limited ability to represent the very-large scale motions (Kunkel and Marusic, 2006) or 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 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 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 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 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 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 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}.~~

~~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

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

~~The dataset for this publication is available upon contacting the corresponding author.~~
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

Smith 1968

$$\sigma = ax^b$$

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.

~~11~~12 Acknowledgements

We would like to thank all members of the fieldwork team including Stephany Paredes-Mesa, Tanvir Mangat and Kira Olander. We would also like to thank Maider Llaguno Munitxa for her help with the mobile tower modeling. We thank ~~LICOR~~LI-COR Biosciences for lending instrumentation used in this campaign. This work was funded by NOAA CPO/AC4, #NA14OAR4310134.

~~12~~13 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, 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, 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, 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 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, 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, 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 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 ~~3~~5. 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	0.54	--	--	--	--
1	2	1.33	40	0	1	150
	3	0.47	0	15	7	13
	1	0.14	--	--	--	--
2	2	0.15	58	49	1	7.5
	3	0.14	72	0	7	5

~~Table 5.~~

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%

Table ~~6~~7. 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

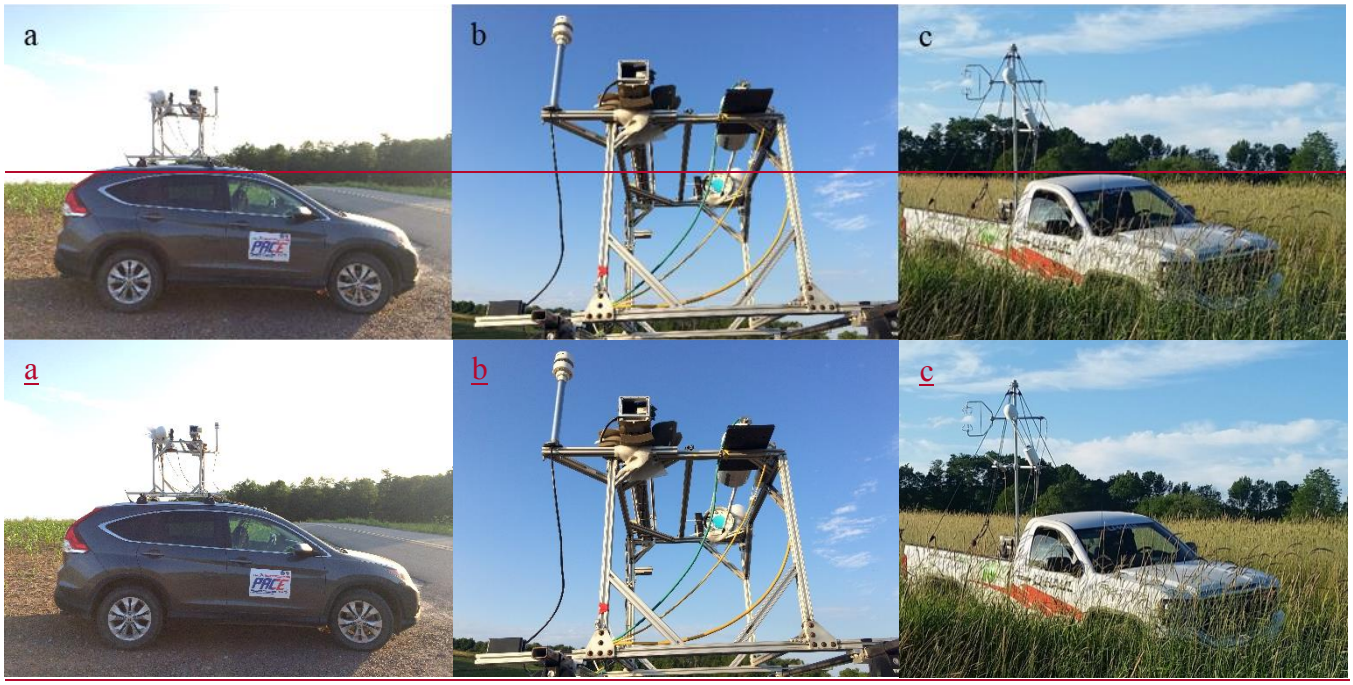


Figure 1. (a) The Princeton atmospheric chemistry experiment (PACE), (b) close-up of PACE roof rack with instrumentation and (c) mobile tower platform.

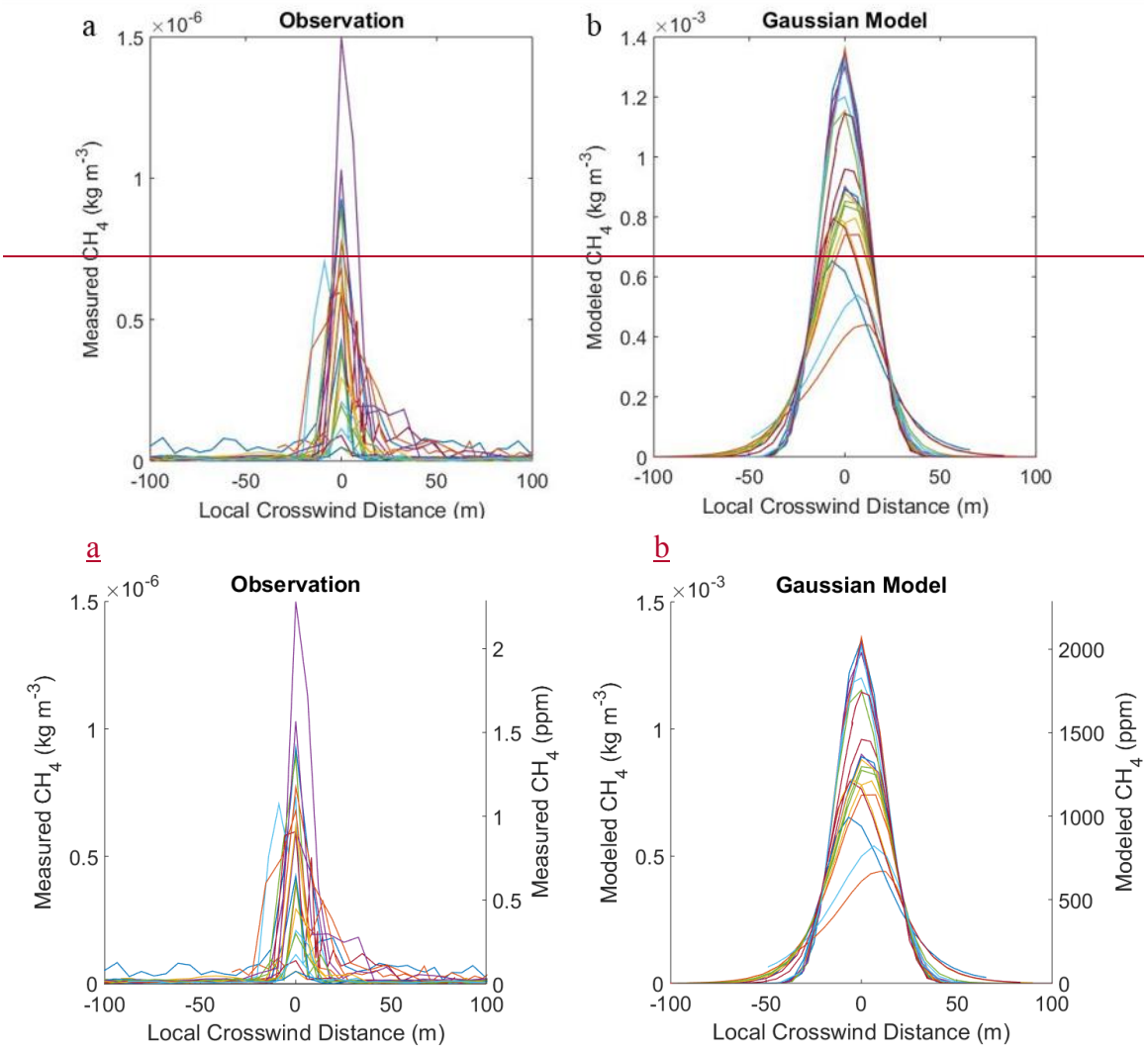


Figure 2. (a) Observations of CH_4 at a site showing multiple downwind transects ~~and~~versus the local crosswind (y) distance. (b) Gaussian outputs along the same downwind transects. Note that the scale difference between the panels is due to the inverse methodology used where the model is set to a reference emission rate (of 1 kg s^{-1}).

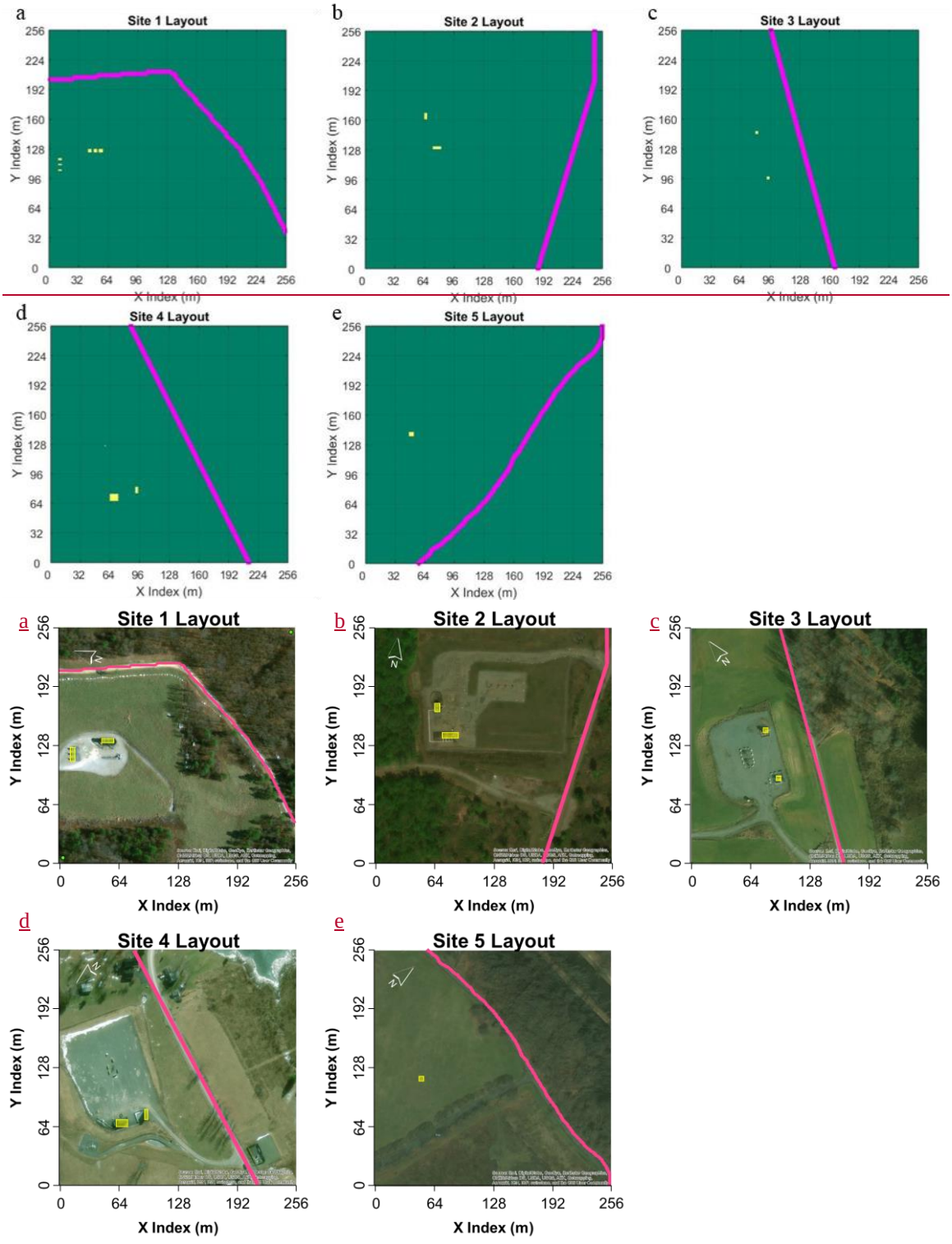

a
Site 1 Layout
b
Site 2 Layout
c
Site 3 Layout
d
Site 4 Layout
e
Site 5 Layout
Y Index (m)
X Index (m)
0
32
64
96
128
160
192
224
256

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.

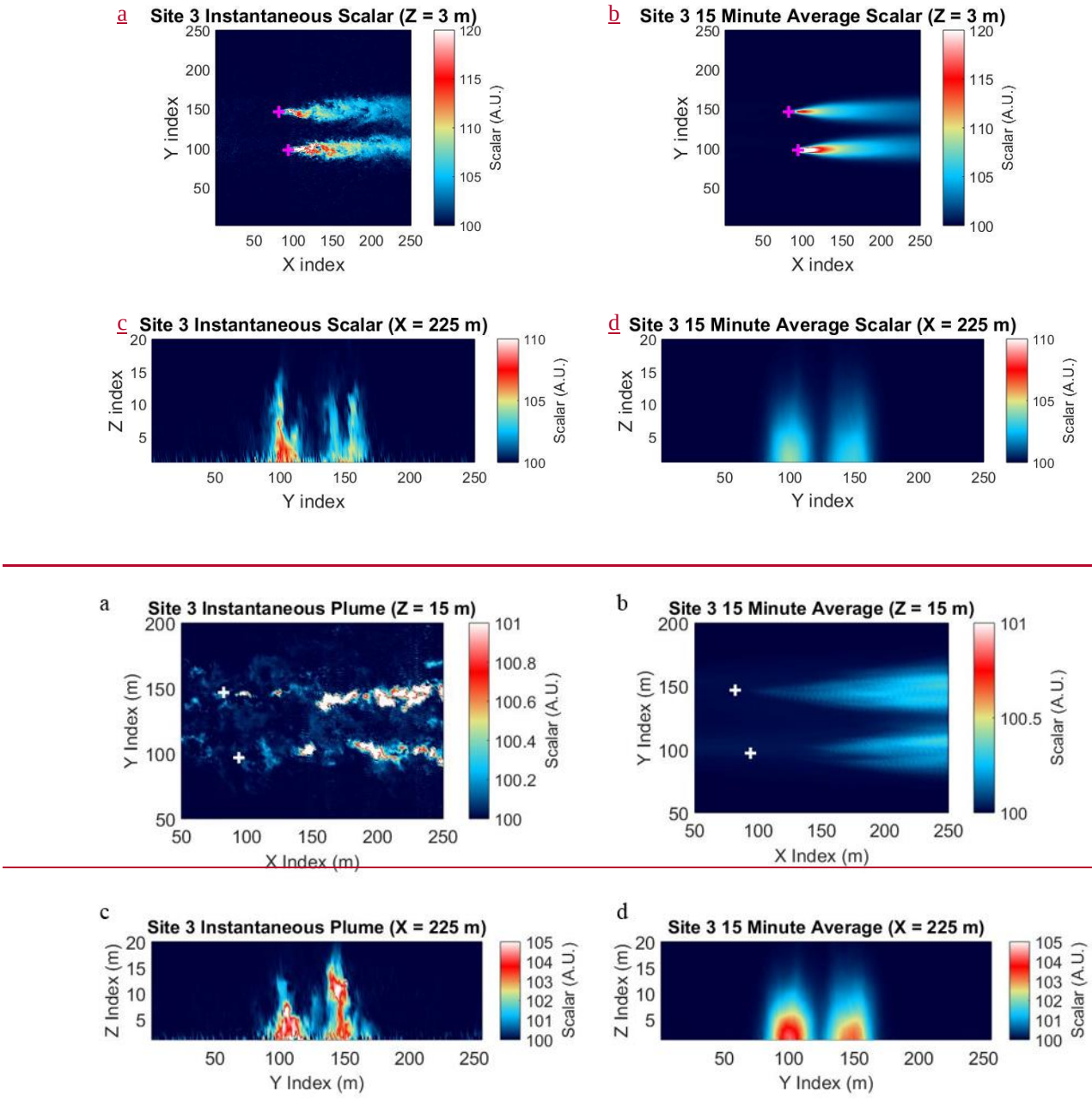
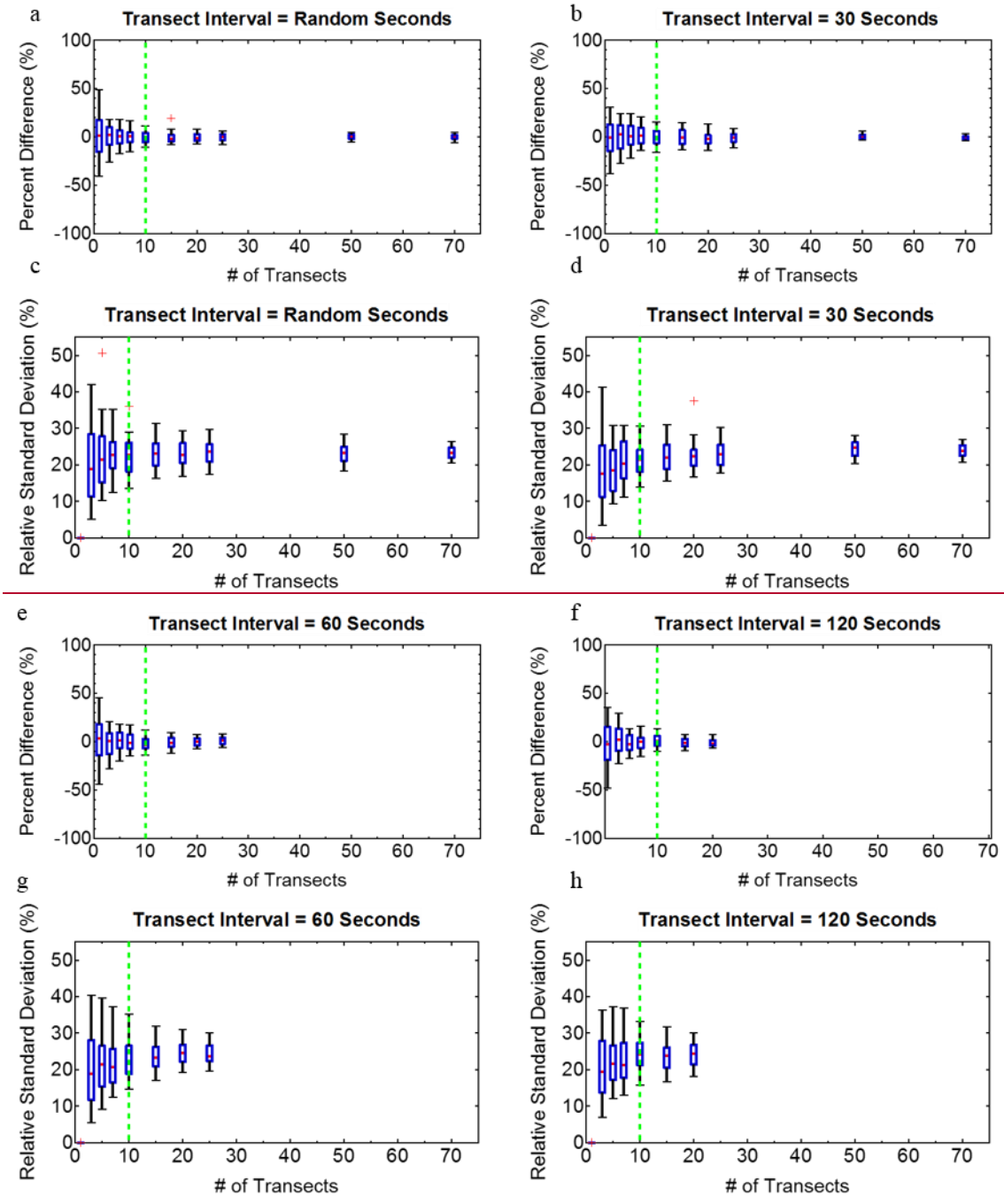


Figure 4. Comparison of instantaneous (a/c) and 15 min averaged (b/d) plume for Site 3~~.~~ from the LES. Panels a and b show scalar with arbitrary units (A.U.) in a x-y cross~~-~~section at an altitude of ~~15~~3 m and panels c and d show a x-z cross~~-~~section at a 225 m downwind distance. The release locations are shown with a ~~white~~magenta marker (+).


a
Transect Interval = Random Seconds
Percent Difference (%)
# of Transects
b
Transect Interval = 30 Seconds
Percent Difference (%)
# of Transects
c
Transect Interval = Random Seconds
Relative Standard Deviation (%)
# of Transects
d
Transect Interval = 30 Seconds
Relative Standard Deviation (%)
# of Transects
e
Transect Interval = 60 Seconds
Percent Difference (%)
# of Transects
f
Transect Interval = 120 Seconds
Percent Difference (%)
# of Transects
g
Transect Interval = 60 Seconds
Relative Standard Deviation (%)
# of Transects
h
Transect Interval = 120 Seconds
Relative Standard Deviation (%)
# of Transects

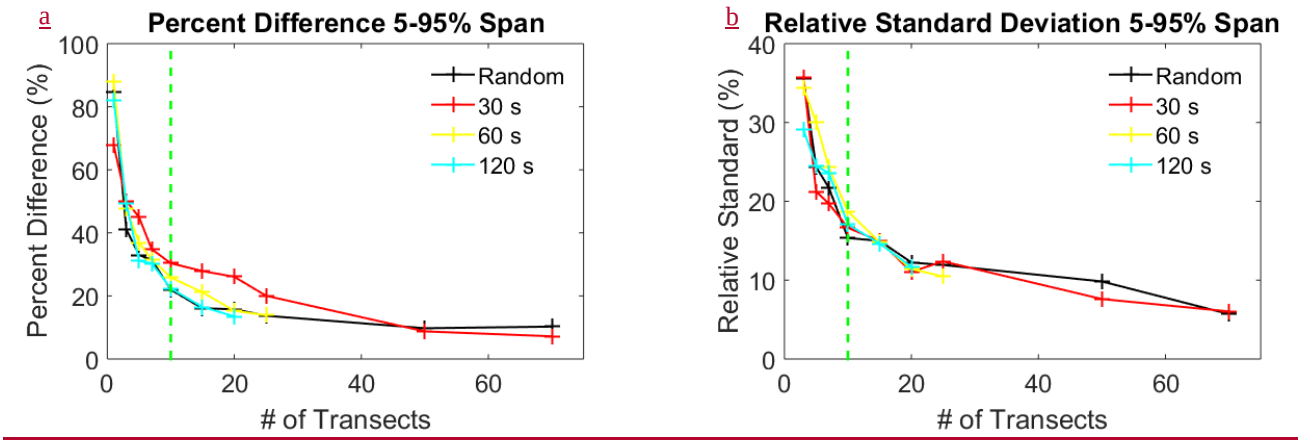
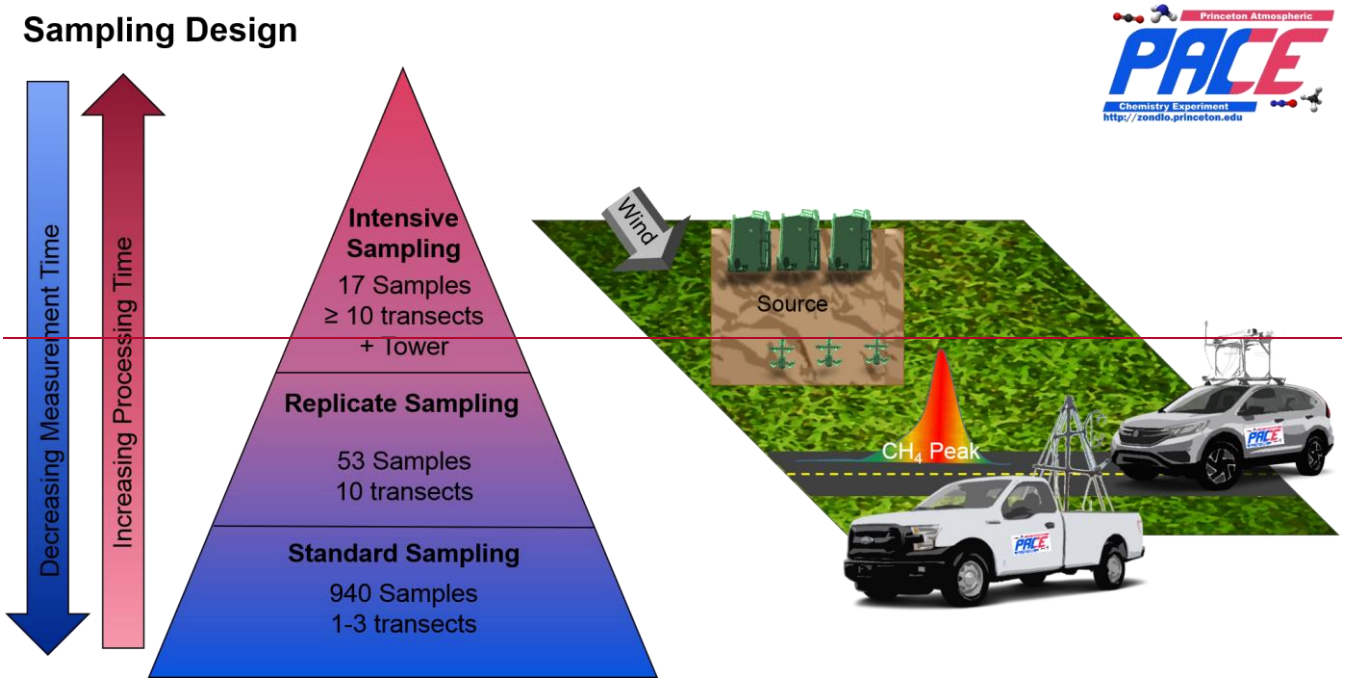


Figure 5. The 5-95% percent difference (a~~,b,e,f~~) between emission retrieval and known simulation emission rate and rsd (~~c,d,g,h~~b) of the emission retrieval using various amounts of transects and random ~~(a,c),~~ 30 second ~~(b,d),~~ 1 minute ~~(e,g)~~ or 2 minute ~~(f,h)~~ transect spacing. ~~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.



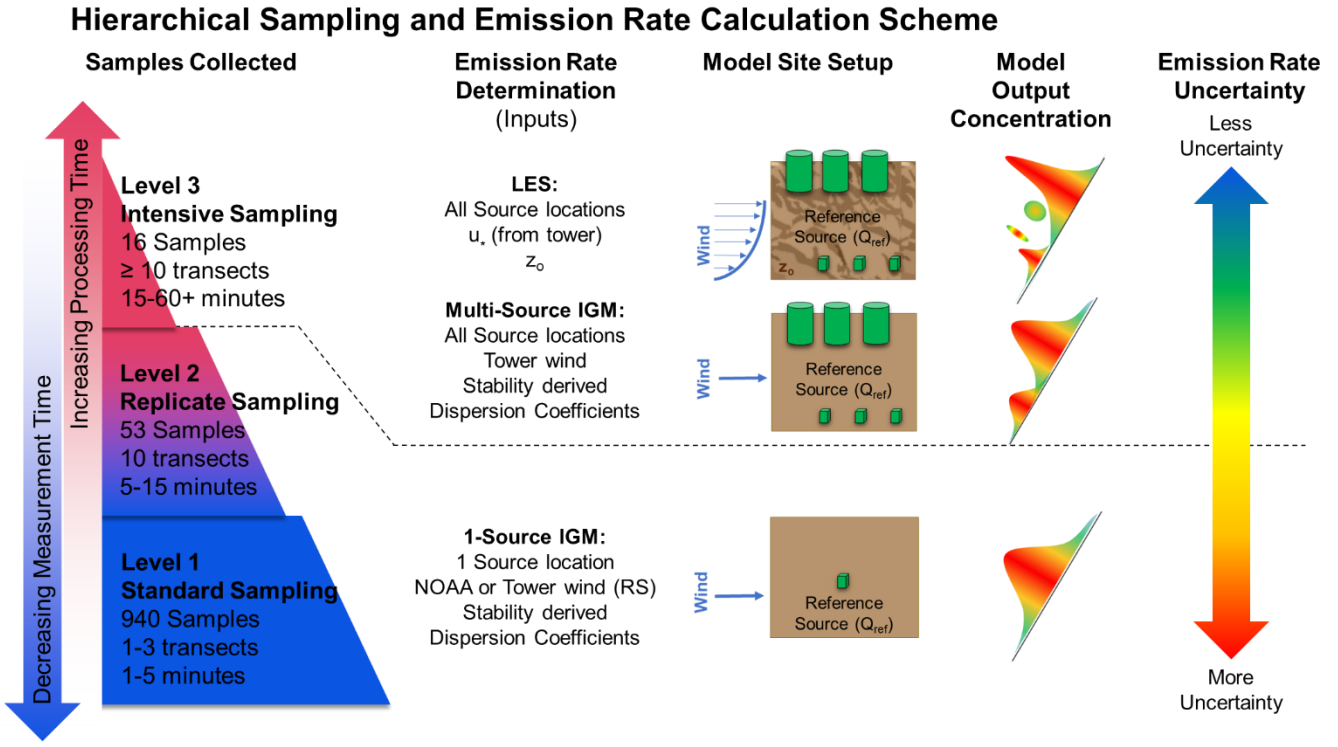
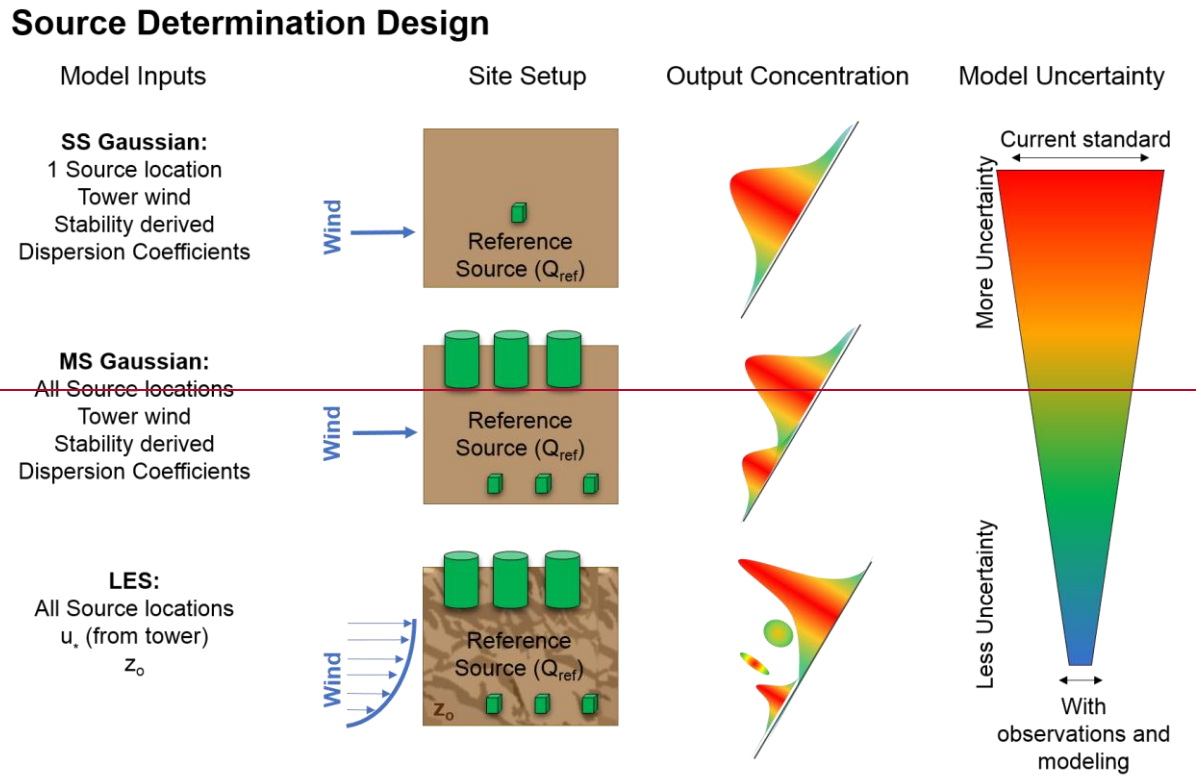
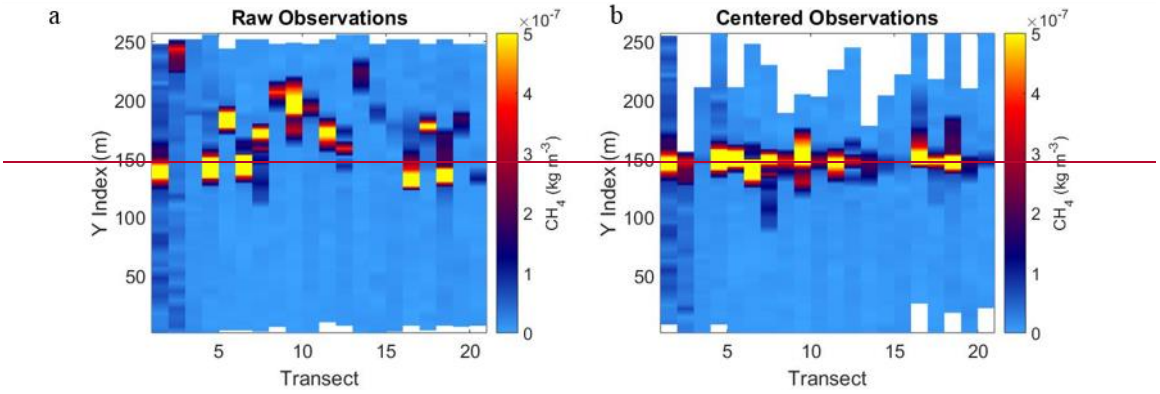


Figure 6. Finalized sampling and emission rate calculation strategy employed in this study showing actual measurements with increasing complexity and decreasing sample size.



~~Figure 7. Finalized source strength determination strategy employed in this study showing~~ as well as models with increasing complexity and decreasing uncertainty. In this schematic u_* is friction velocity and z_o is terrain roughness length. 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).


Site 3
a
Raw Observations
b
Centered Observations
Y Index (m)
Transect
CH4 (kg m-3)
×10-7

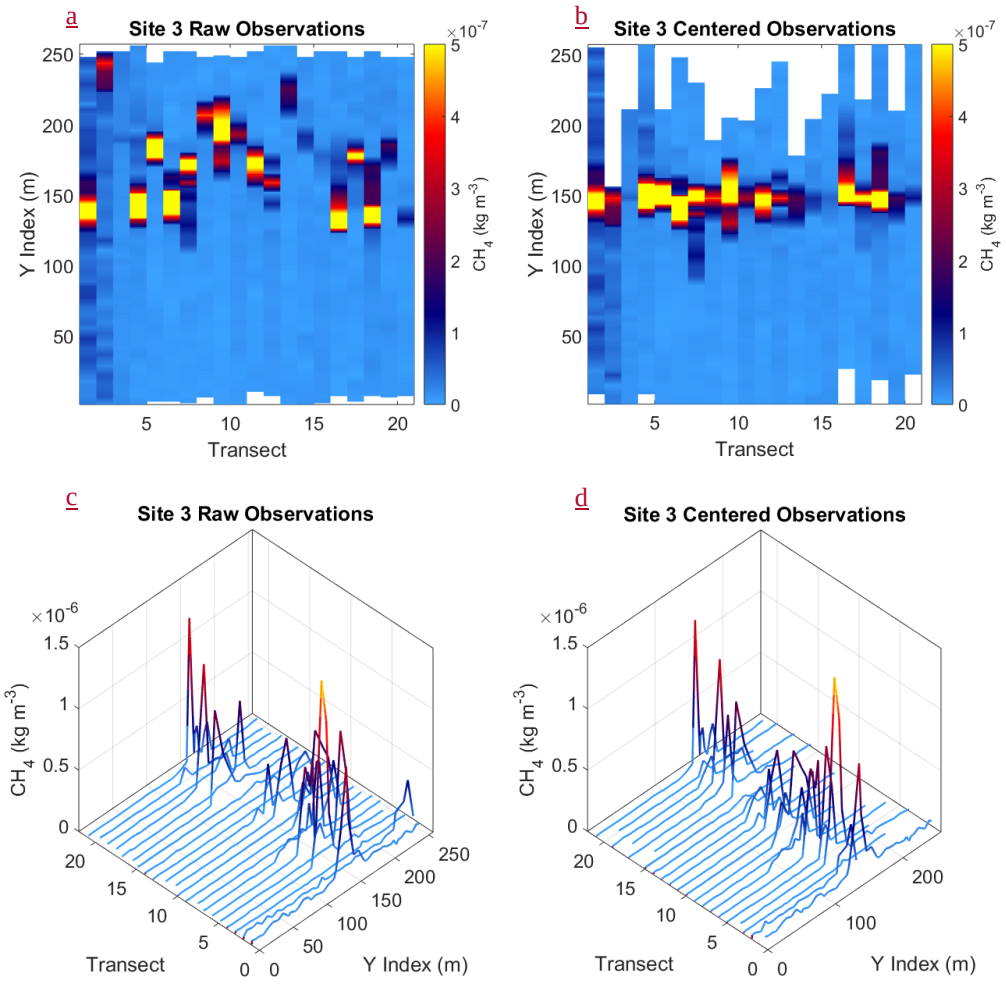


Figure ~~8~~7. (a) Site 3 raw observations as indexed to the LES domain and (b) after they have been aligned with the peak of the LES plume. Waterfall plots of the same data for (a) raw observations and (b) centered observations.

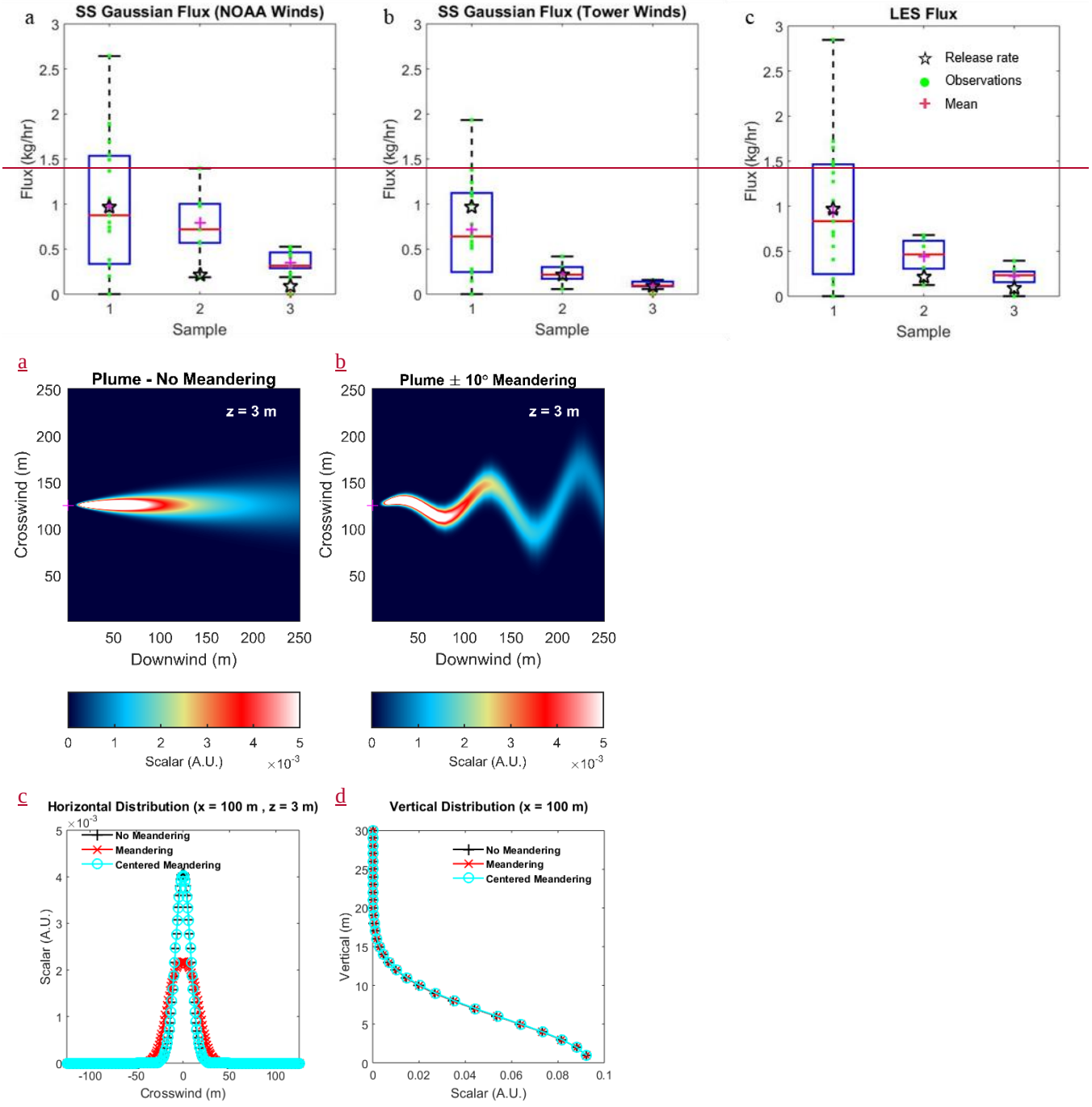


Figure 8. A theoretical site with an additional plume meandering scale added shown in a top-down view at 3 m of (a) Gaussian with no meandering, (b) instantaneous Gaussian with 10° meandering. The comparison of (c) horizontal

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.

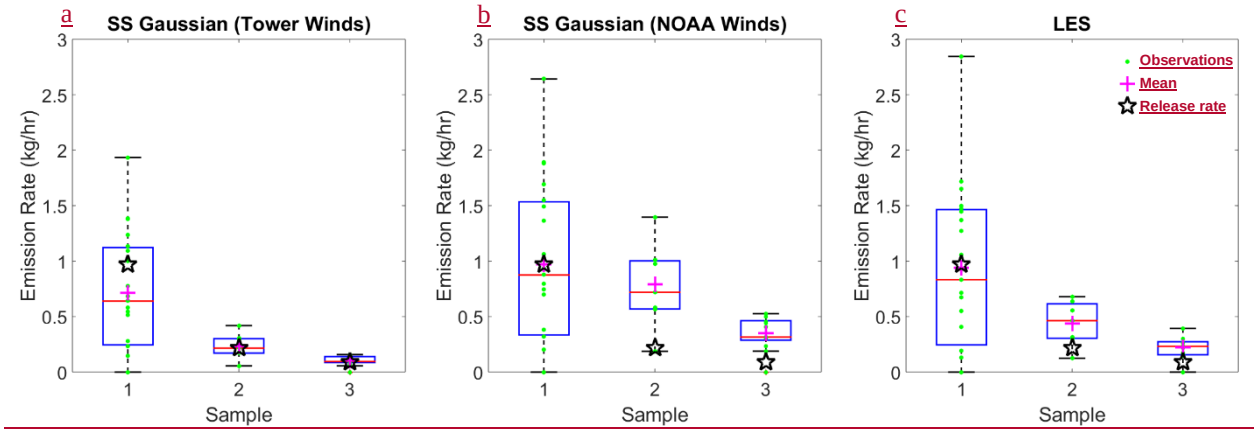
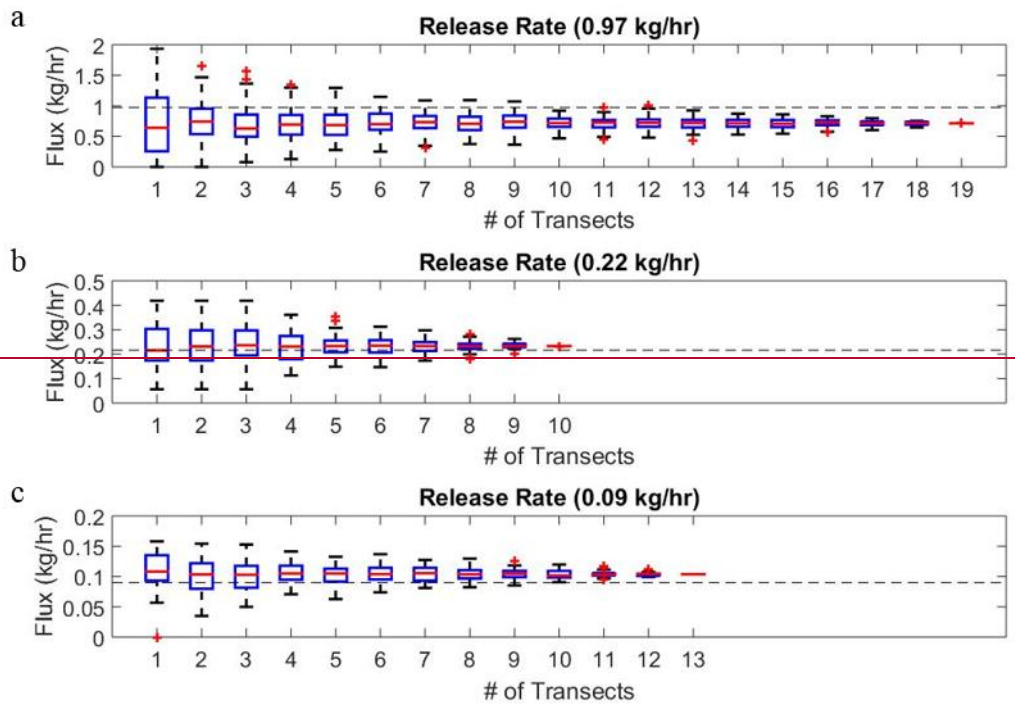


Figure 9. Results from three controlled release experiments for (a) the Gaussian approach using NOAA winds, (b) the Gaussian approach using tower measured winds and (c) from the LES. Box and whiskers plots show the 50% percentile (red), 25 and 75% percentile (blue) and minimum and maximum values (black). 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)
Flux (kg/hr)
# of Transects
b
Release Rate (0.22 kg/hr)
Flux (kg/hr)
# of Transects
c
Release Rate (0.09 kg/hr)
Flux (kg/hr)
# of Transects

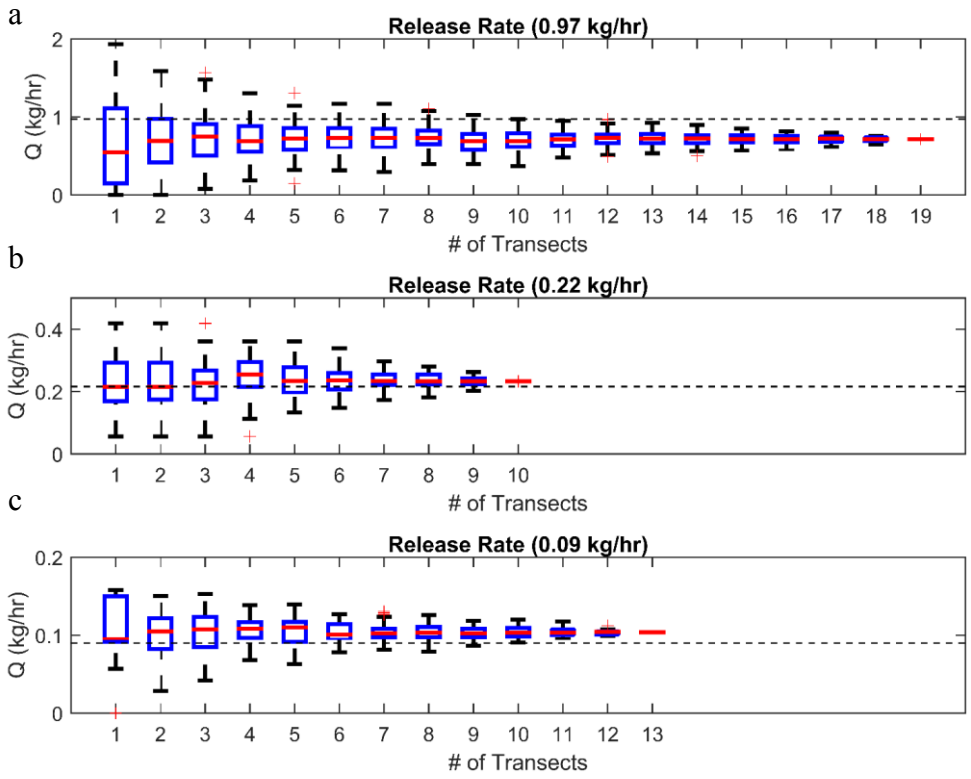
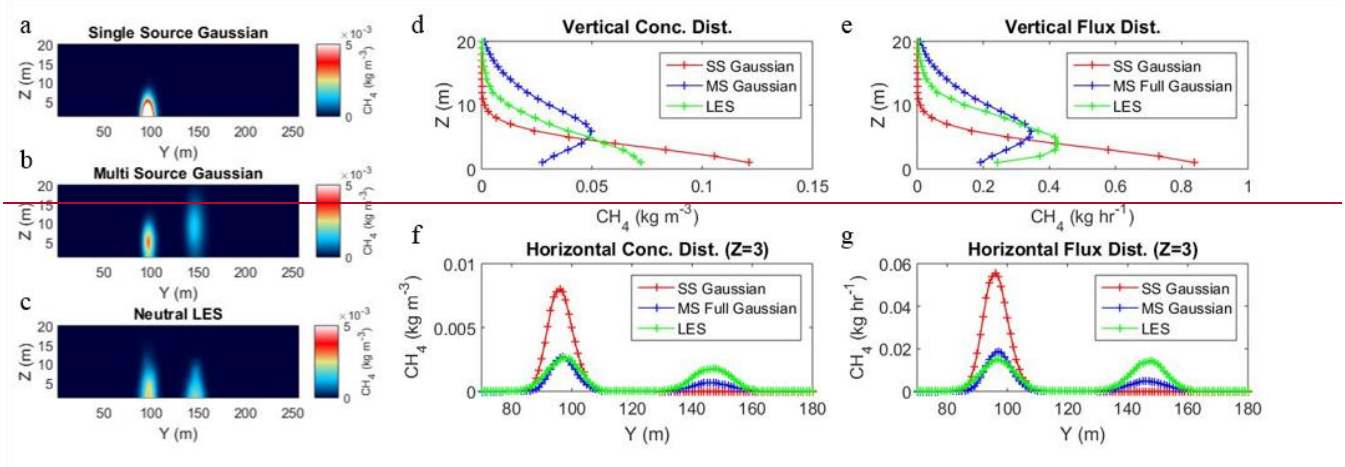


Figure 10. A comparison of the convergence of the plume rate inferred from the IGM by averaging randomly selected transects for (a) the 0.97 kg hr^{-1} release rate, (b) the 0.22 kg hr^{-1} release rate and (c) the 0.09 kg hr^{-1} release rate. The actual release rate is shown as a dashed black line. 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.



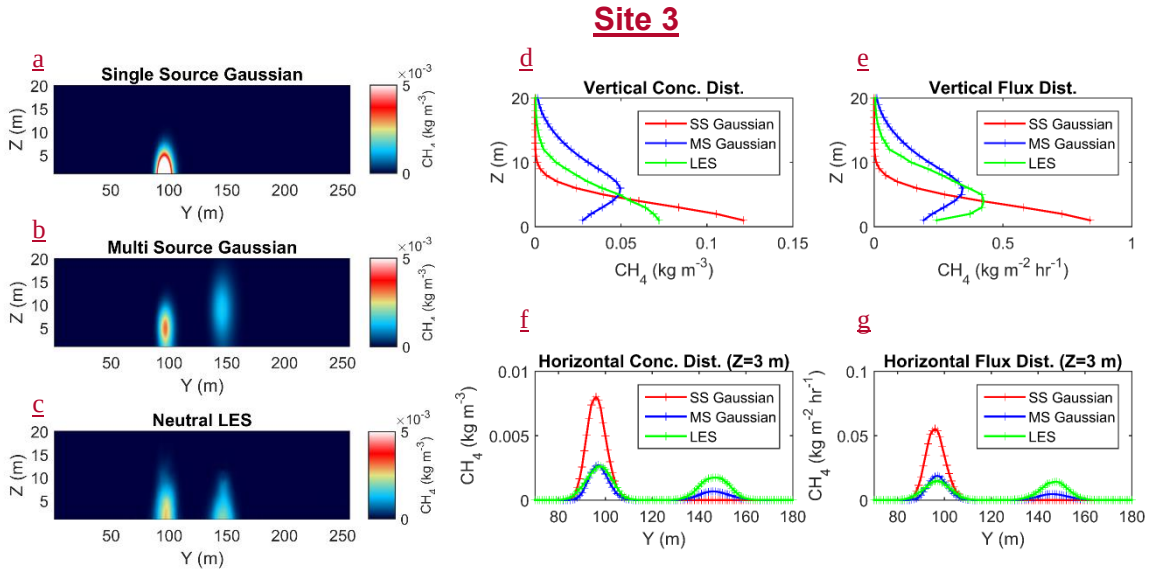
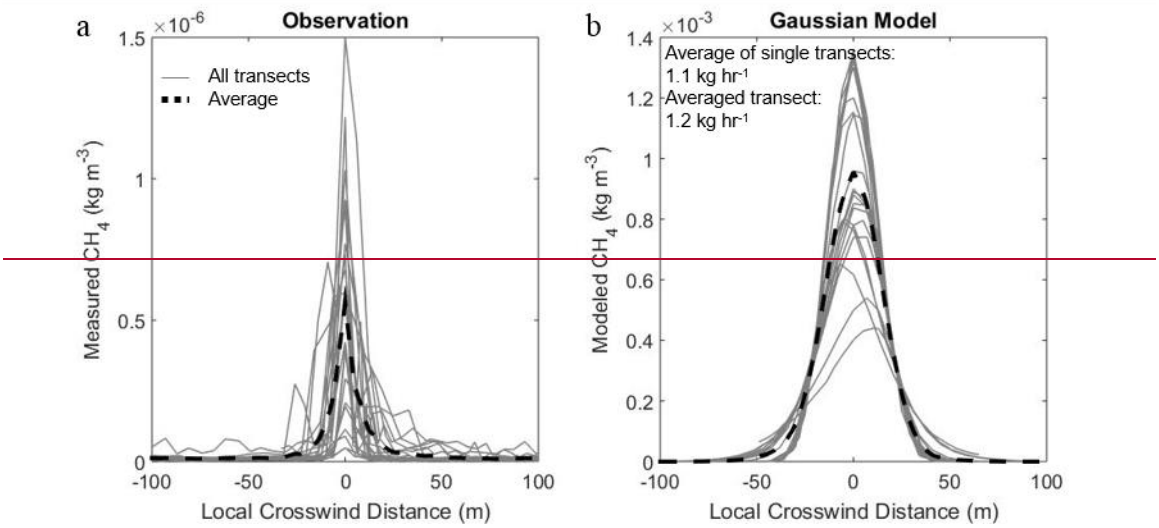


Figure 11. Comparison of three scenarios for Site 3 showing images through the downwind road plane (~30m downwind as shown in Fig. 3) of (a) single source Gaussian, (b) multi-source Gaussian and (c) averaged LES. The comparison of vertical distributions of (d) concentrations and (e) fluxes and of the horizontal distributions of (f) concentrations and (g) fluxes are also shown. The vertical LES flux corresponds to resolved fluxes only.



~~Figure 12. (a) Observations of CH_4 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.~~

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.

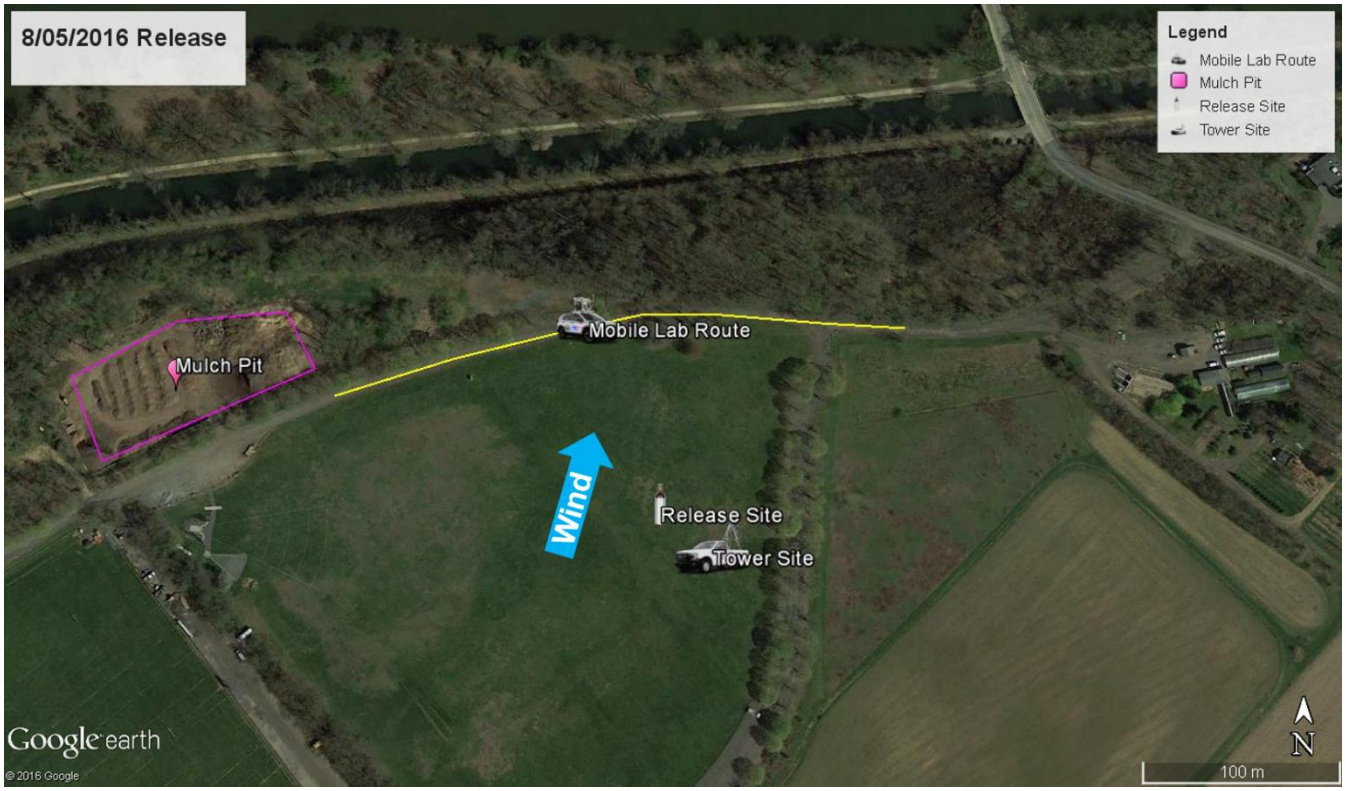


Figure S1. Controlled release site schematic.

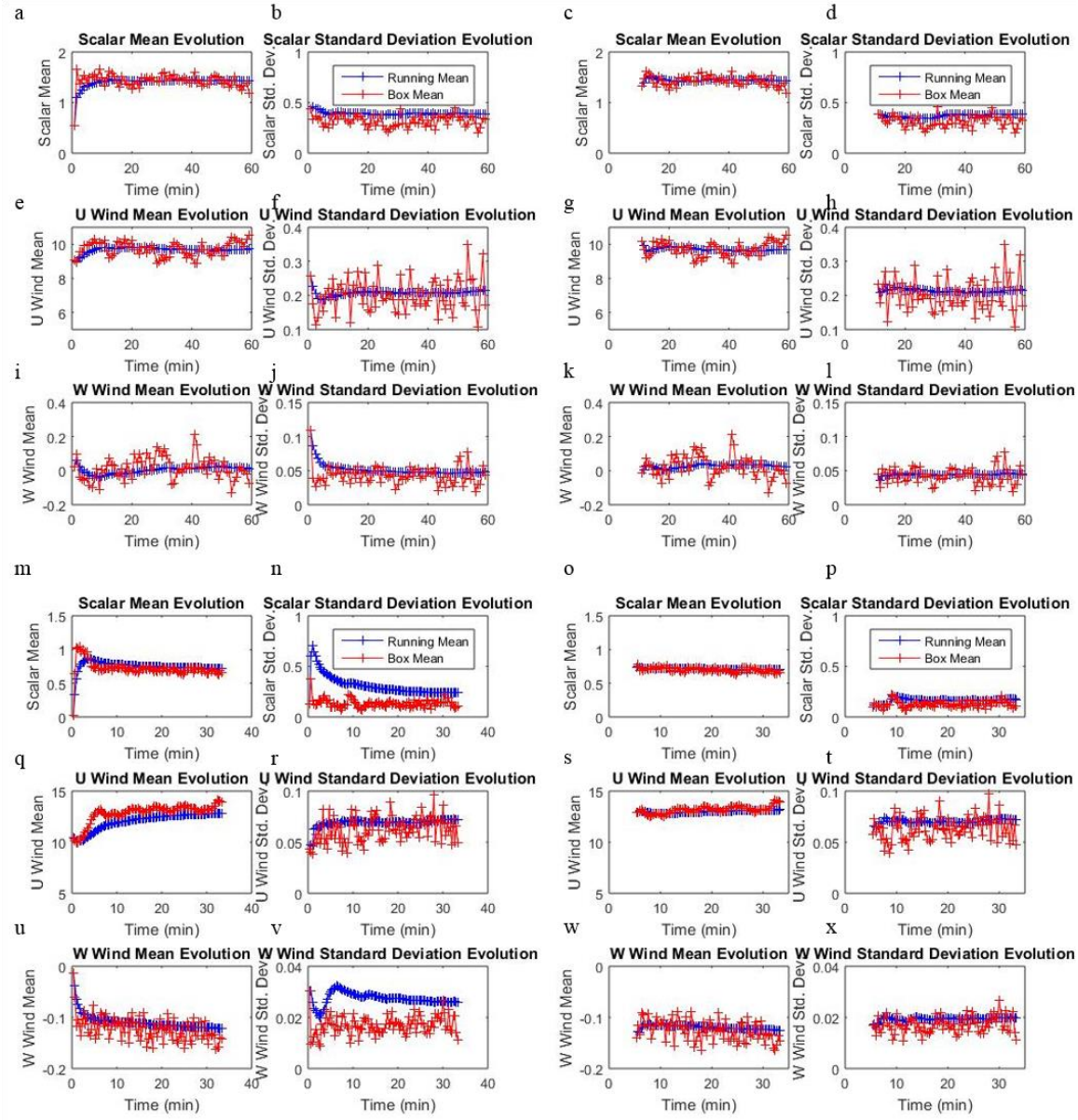
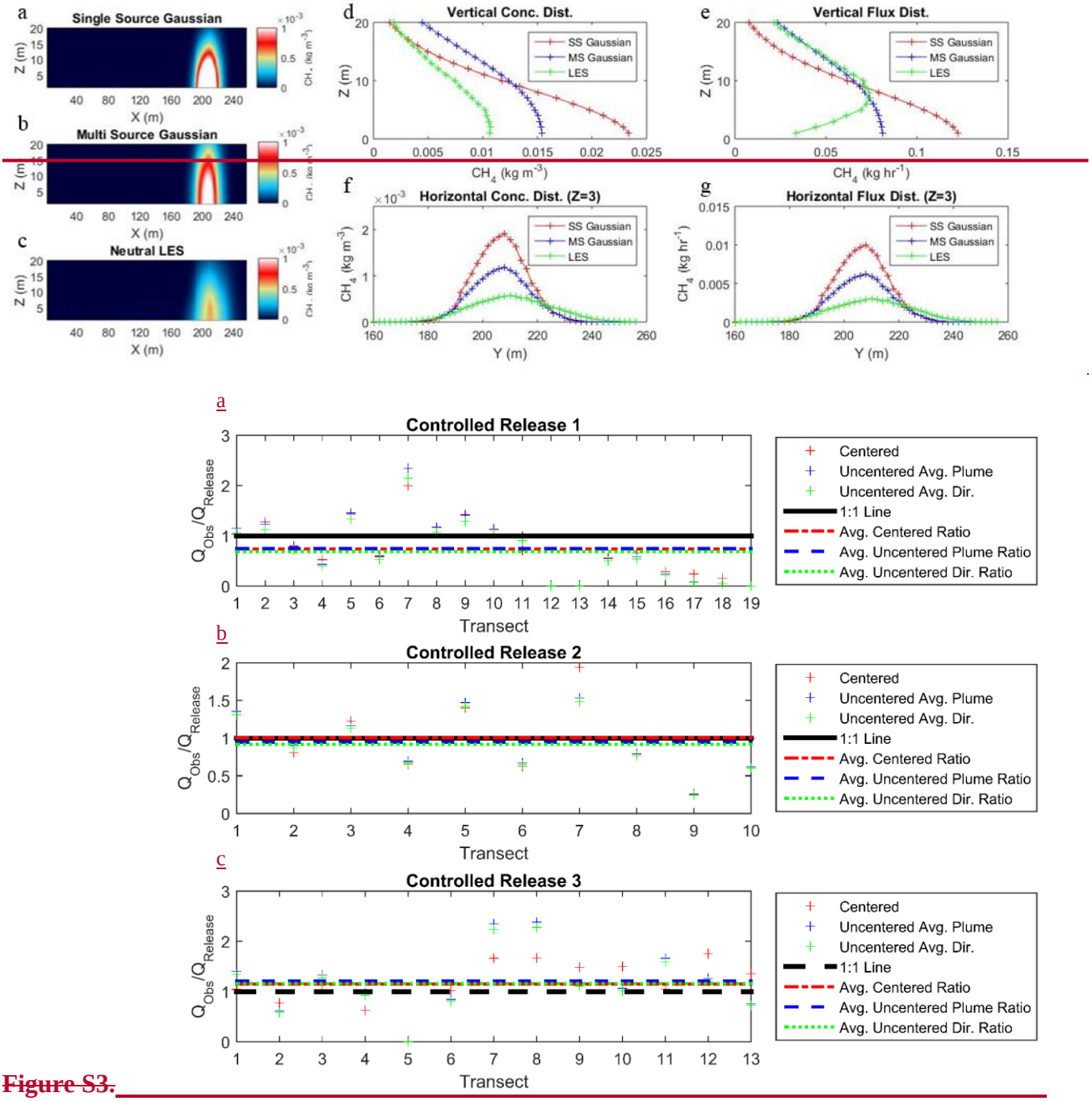


Figure S2. Comparison of box-means or standard deviations (red) and running means or standard deviations (blue) for (a-l) Site 1 and (m-x) Site 2. Left-hand panels show both sites from start-up and right-hand panels are screened to show results after a steady-state has been achieved. The running mean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



~~**Figure S3.**~~ **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.).**

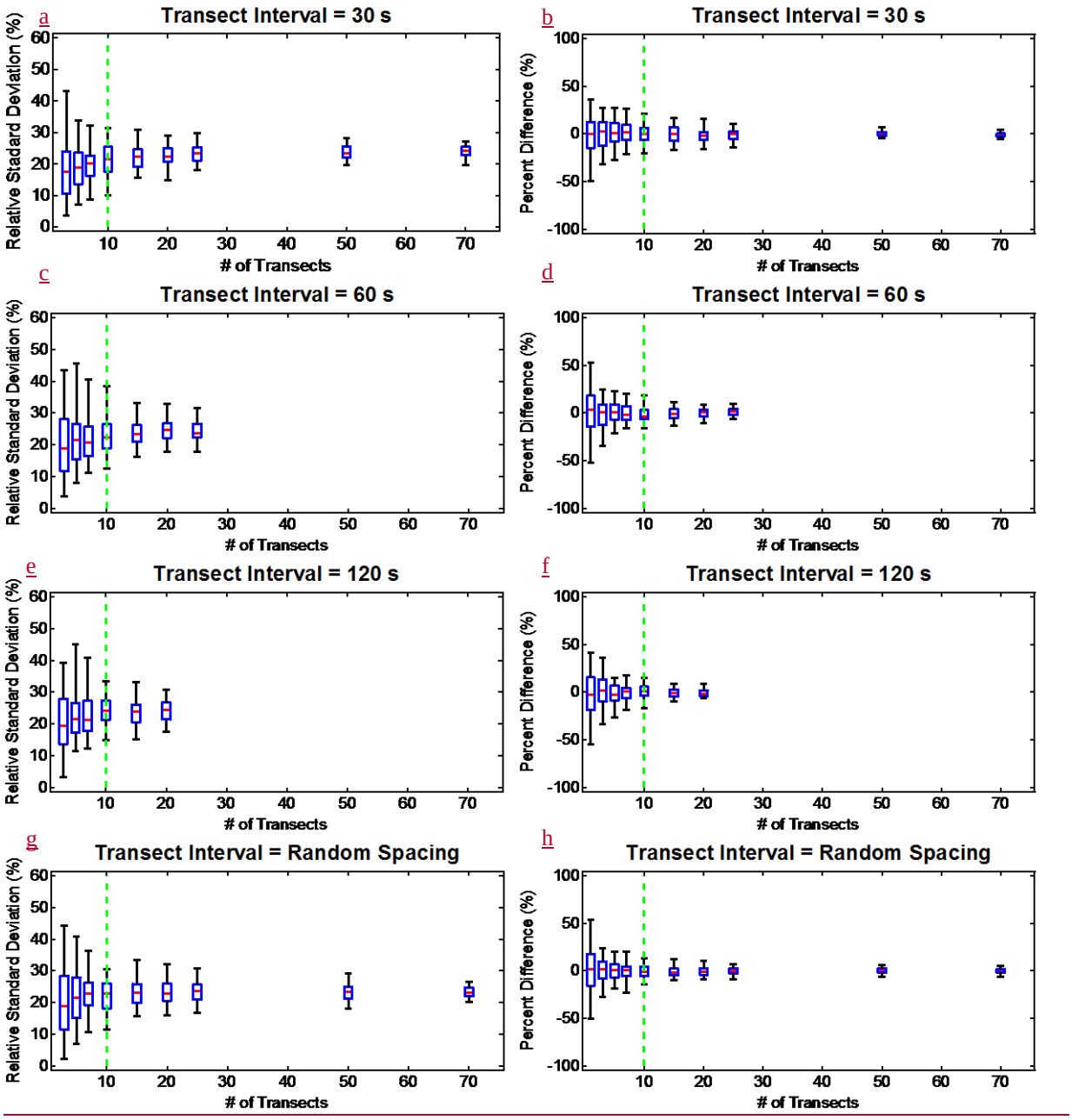


Figure S4. The rsd (a,c,e,g) of the emission retrieval using and the percent difference (b,d,f,h) between emission retrieval and known simulation emission rate using various amounts of transects and 30 second (a,b), 1 minute (c,d), 2 minute (e,f) or random (g,h) transect spacing. Box and whiskers plots 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.

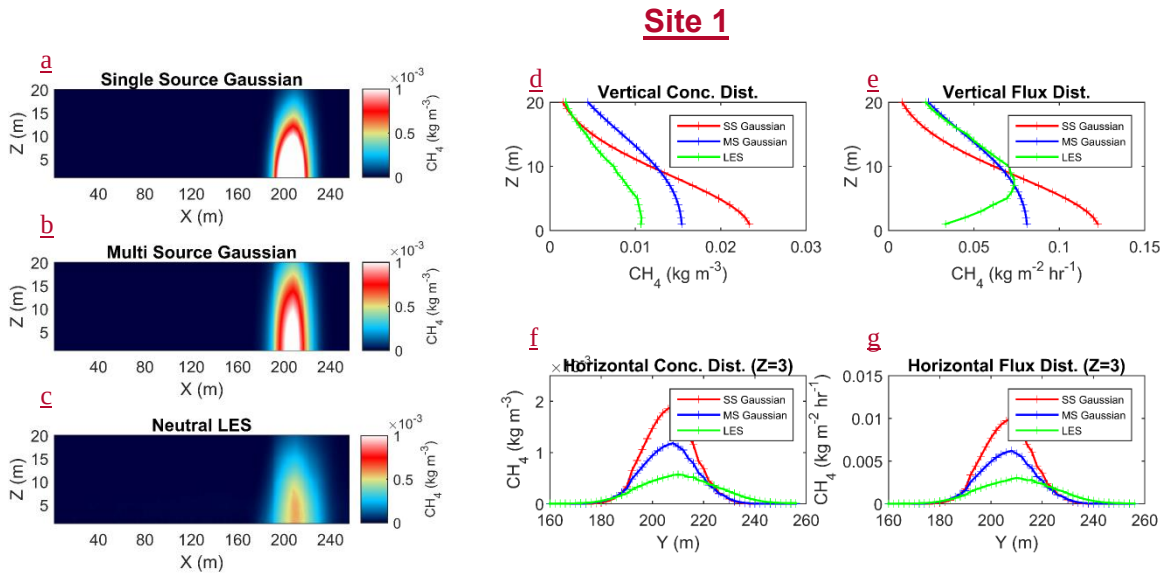
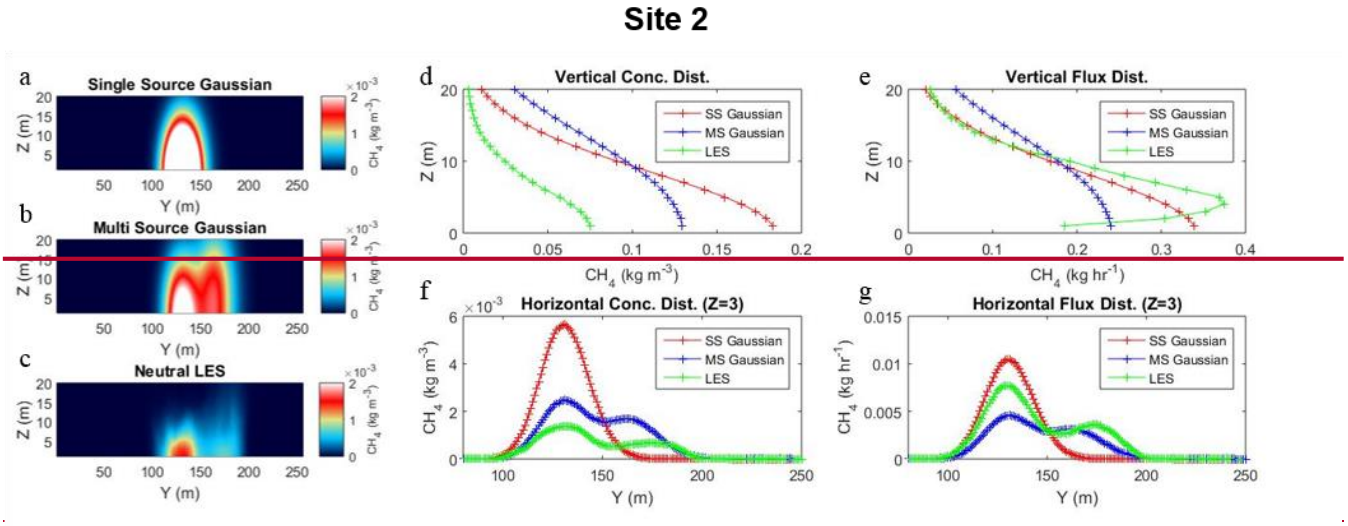


Figure S5. Comparison of three scenarios for Site 1 showing images of (a) single Gaussian, (b) multi-source Gaussian and (c) averaged LES. The comparison of vertical distributions (d) concentrations and (e) fluxes, and horizontal distributions of (f) concentrations and (g) fluxes are also shown.



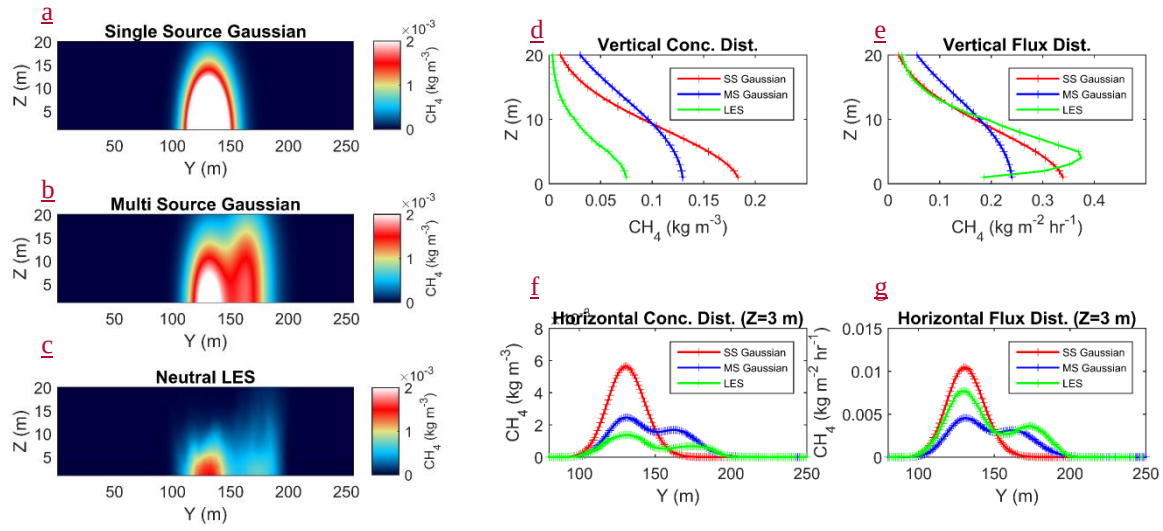
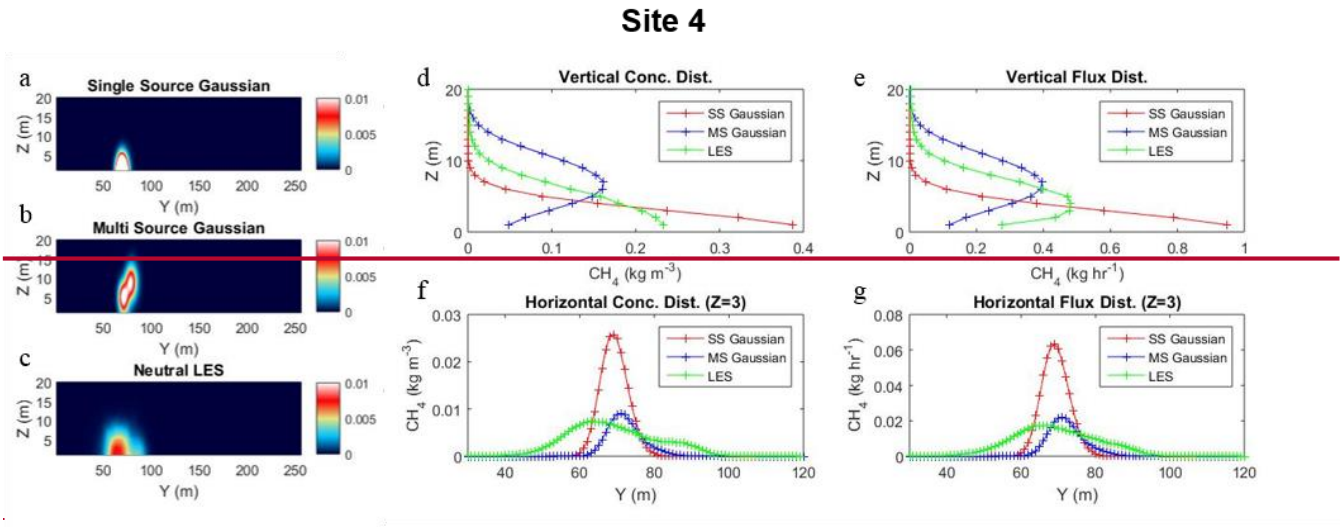


Figure ~~S4~~S6. Comparison of three scenarios for Site 2 showing images of (a) single Gaussian, (b) multi-source Gaussian and (c) averaged LES. The comparison of vertical distributions of (d) concentrations and (e) fluxes, and horizontal distributions of (f) concentrations and (g) fluxes are also shown.



Site 4

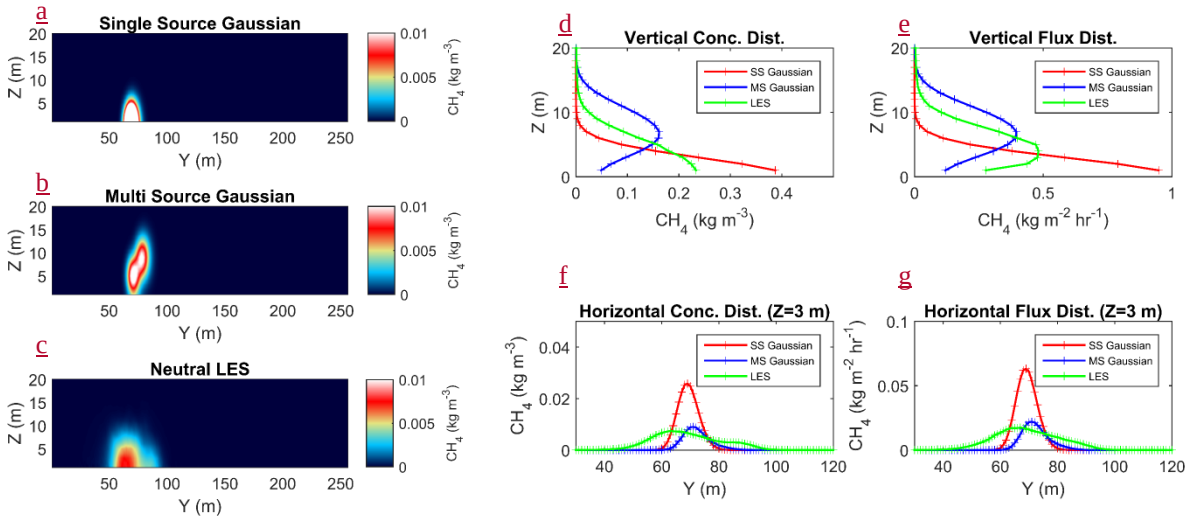
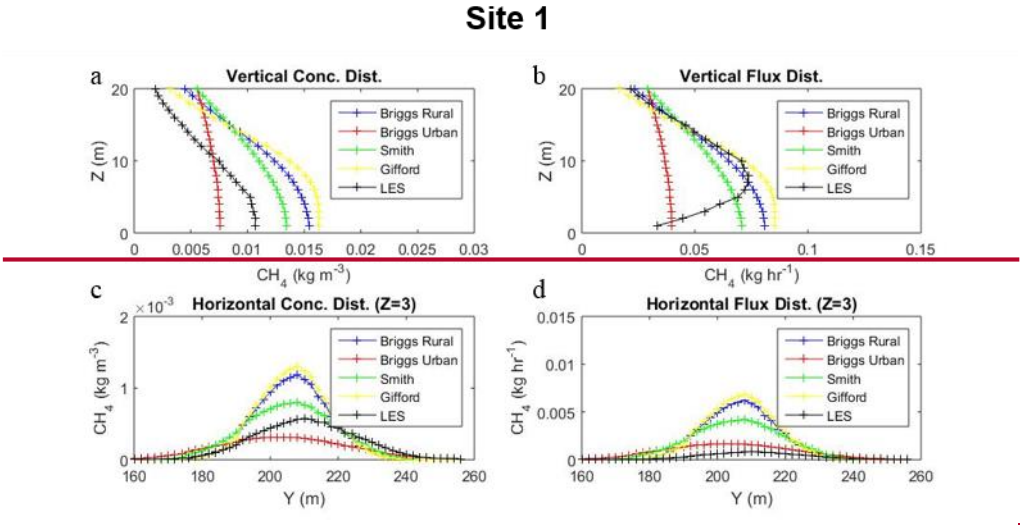


Figure ~~S5~~S7. Comparison of three scenarios for Site 4 showing images of (a) single Gaussian, (b) multi-source Gaussian and (c) averaged LES. The comparison of vertical distributions of (d) concentrations and (e) fluxes, and horizontal distributions of (f) concentrations and (g) fluxes are also shown.



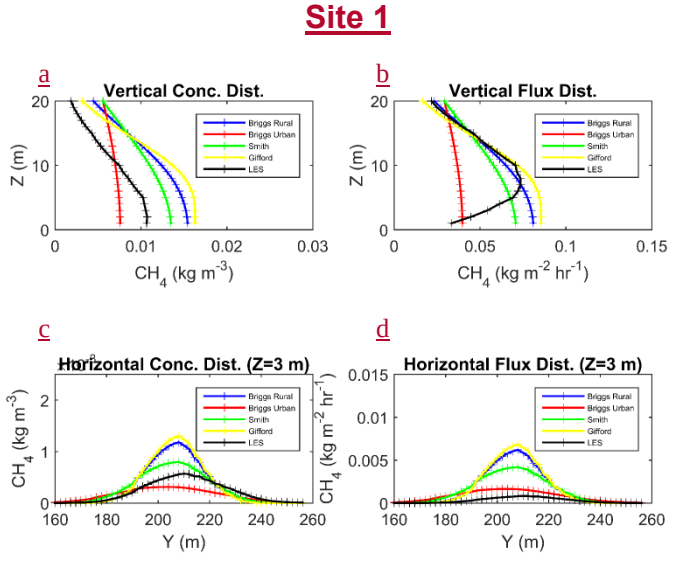


Figure ~~S6~~S8. Comparison of scenarios using different diffusion models for Site 1 showing images of the comparison of vertical distributions of (a) concentration, (b) flux and of the horizontal distributions of (c) concentration and (d) flux.

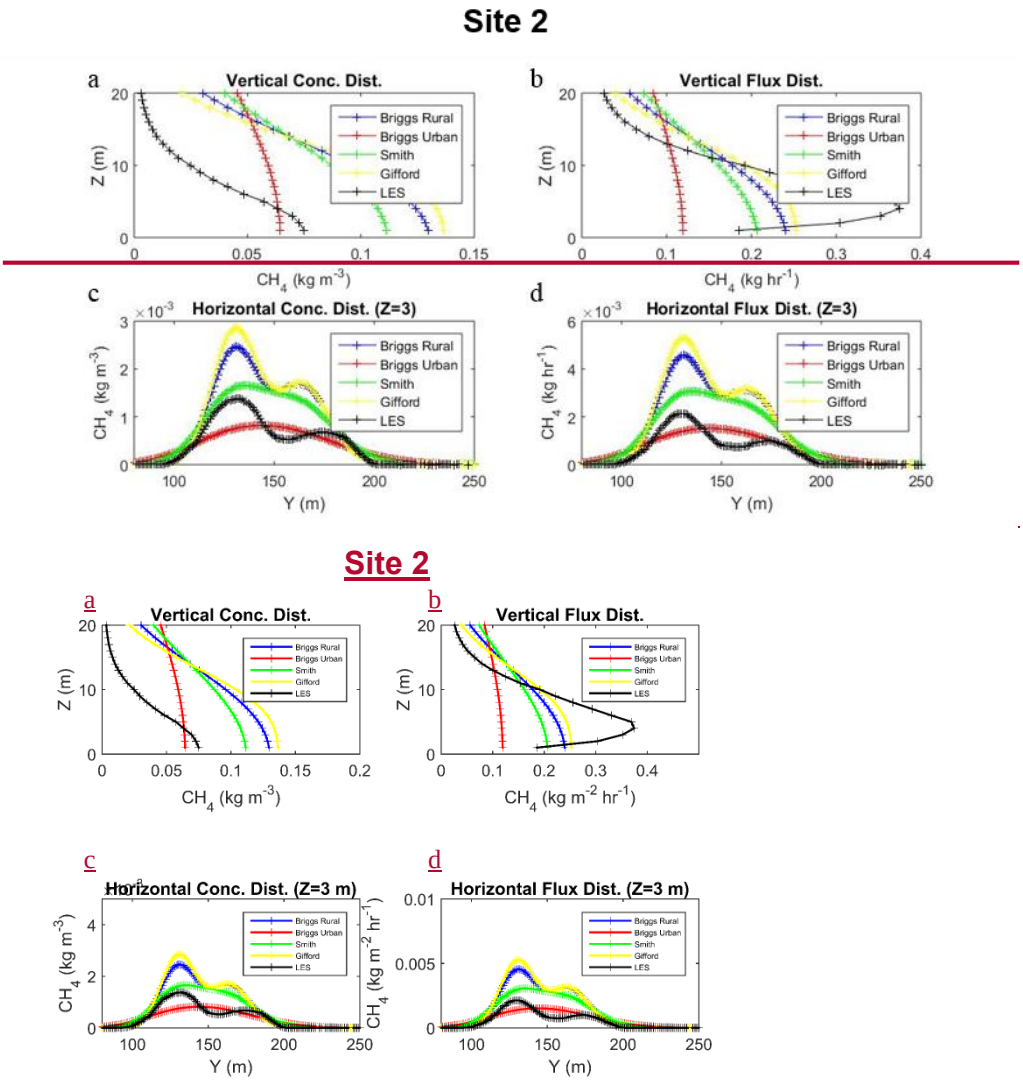


Figure ~~S7~~S9. Comparison of scenarios using different diffusion models for Site 2 showing images of the comparison of vertical distributions of (a) concentration, (b) flux and of the horizontal distributions of (c) concentration and (d) flux.

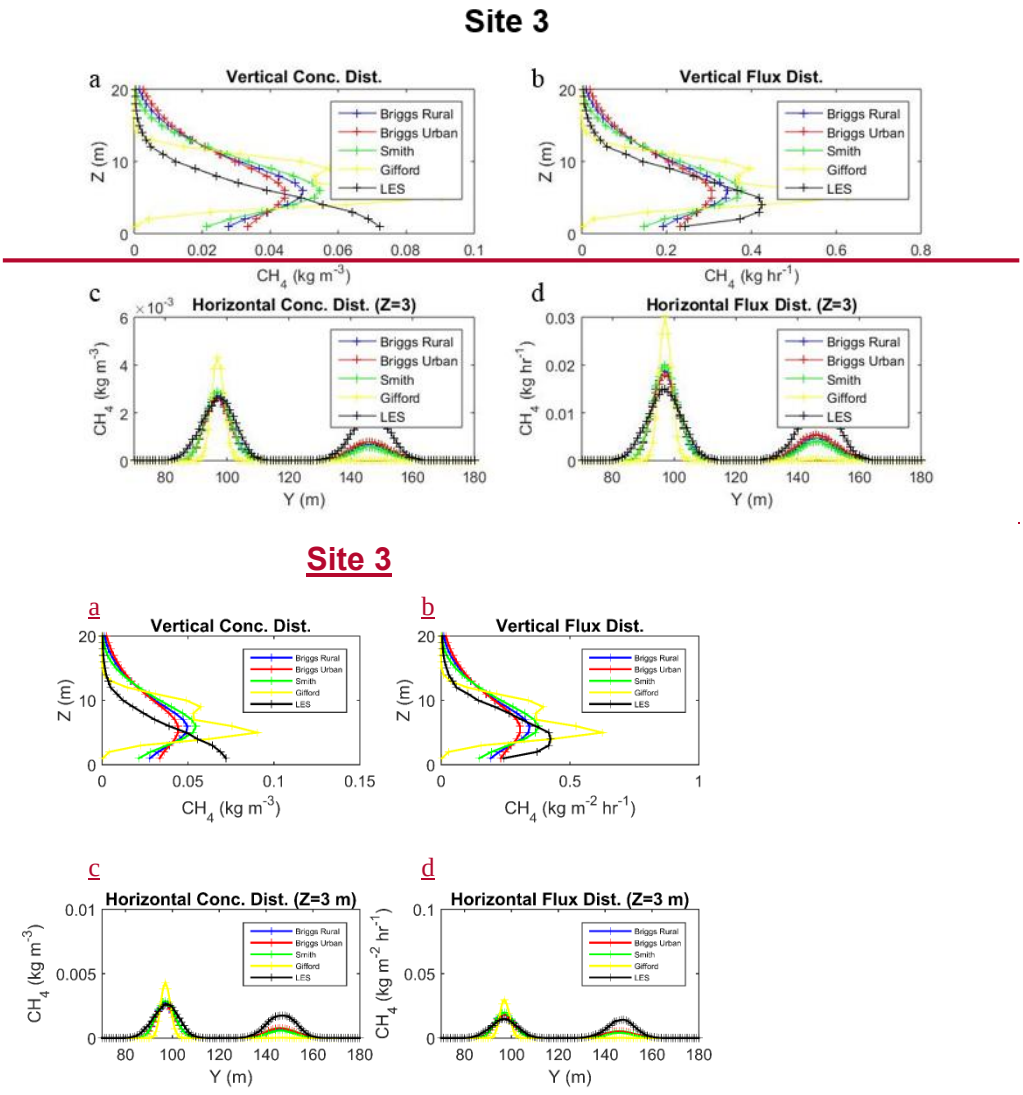
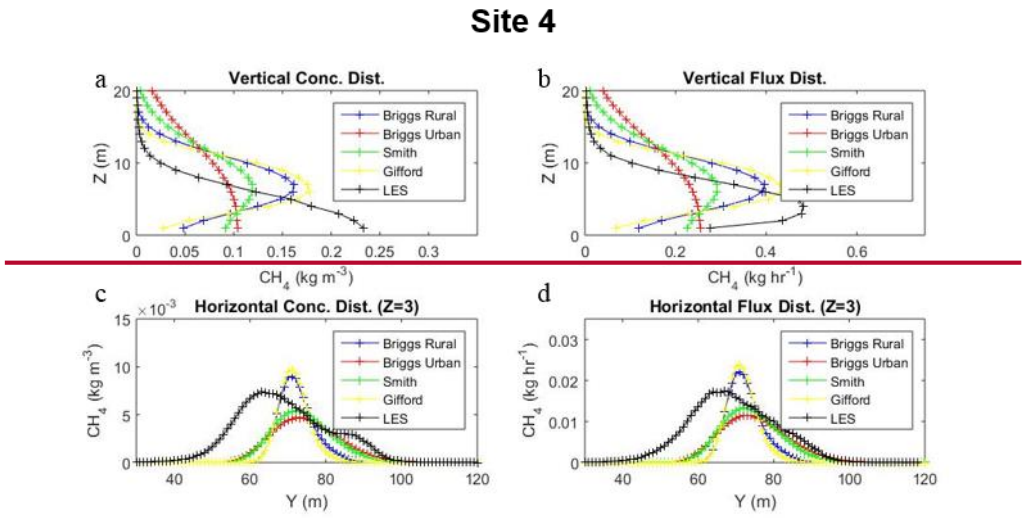


Figure ~~S8~~S10. Comparison of scenarios using different diffusion models for Site 3 showing images of the comparison of vertical distributions of (a) concentration, (b) flux and of the horizontal distributions of (c) concentration and (d) flux.



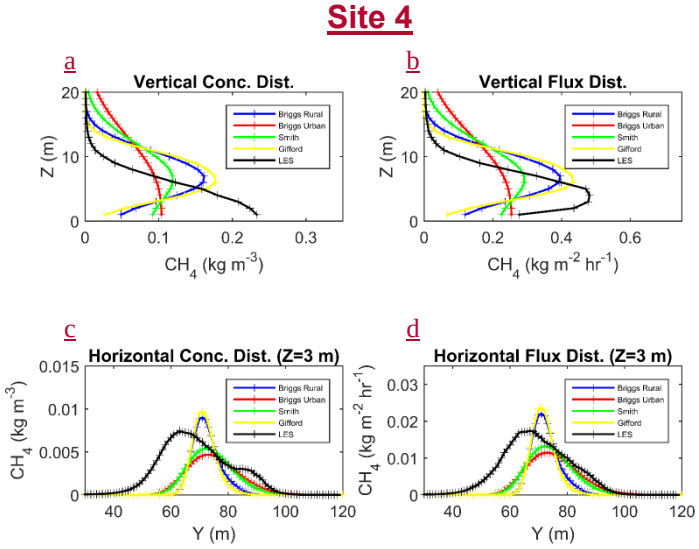
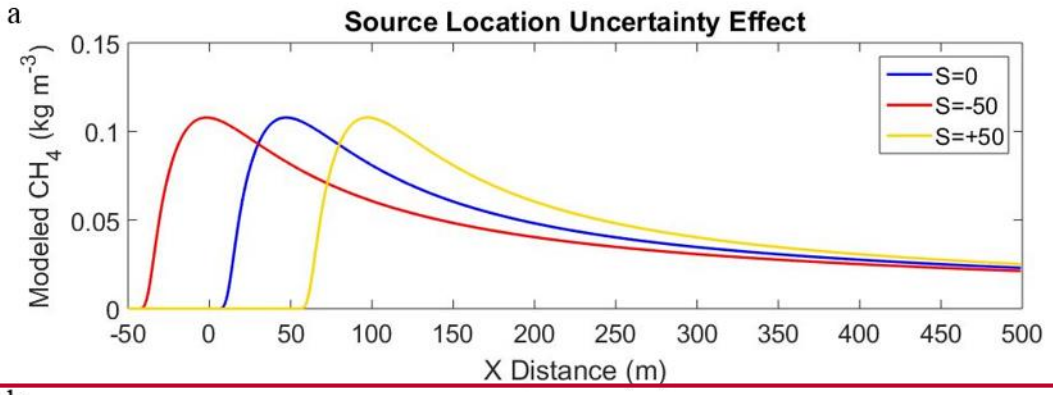
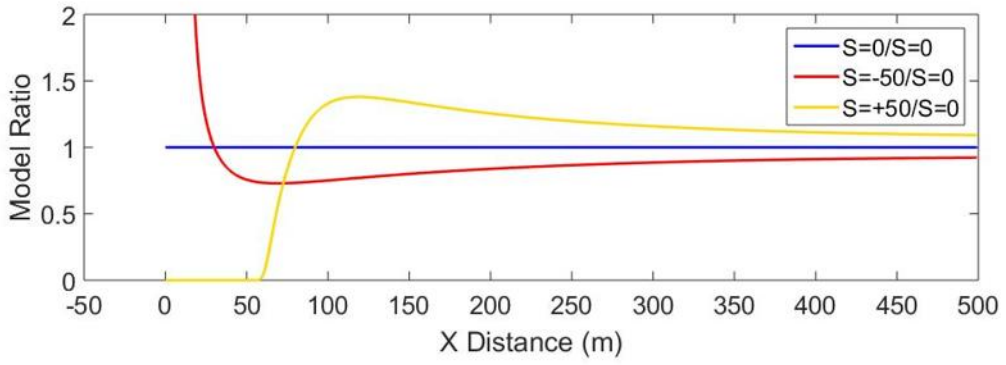


Figure ~~S9~~S11. Comparison of scenarios using different diffusion models for Site 4 showing images of the comparison of vertical distributions of (a) concentration, (b) flux and of the horizontal distributions of (c) concentration and (d) flux.

a


Source Location Uncertainty Effect
Modeled CH_4 (kg m^{-3})
0.15
0.1
0.05
0
-50
0
50
100
150
200
250
300
350
400
450
500
X Distance (m)
S=0
S=-50
S=+50

b


Model Ratio
2
1.5
1
0.5
0
-50
0
50
100
150
200
250
300
350
400
450
500
X Distance (m)
S=0/S=0
S=-50/S=0
S=+50/S=0

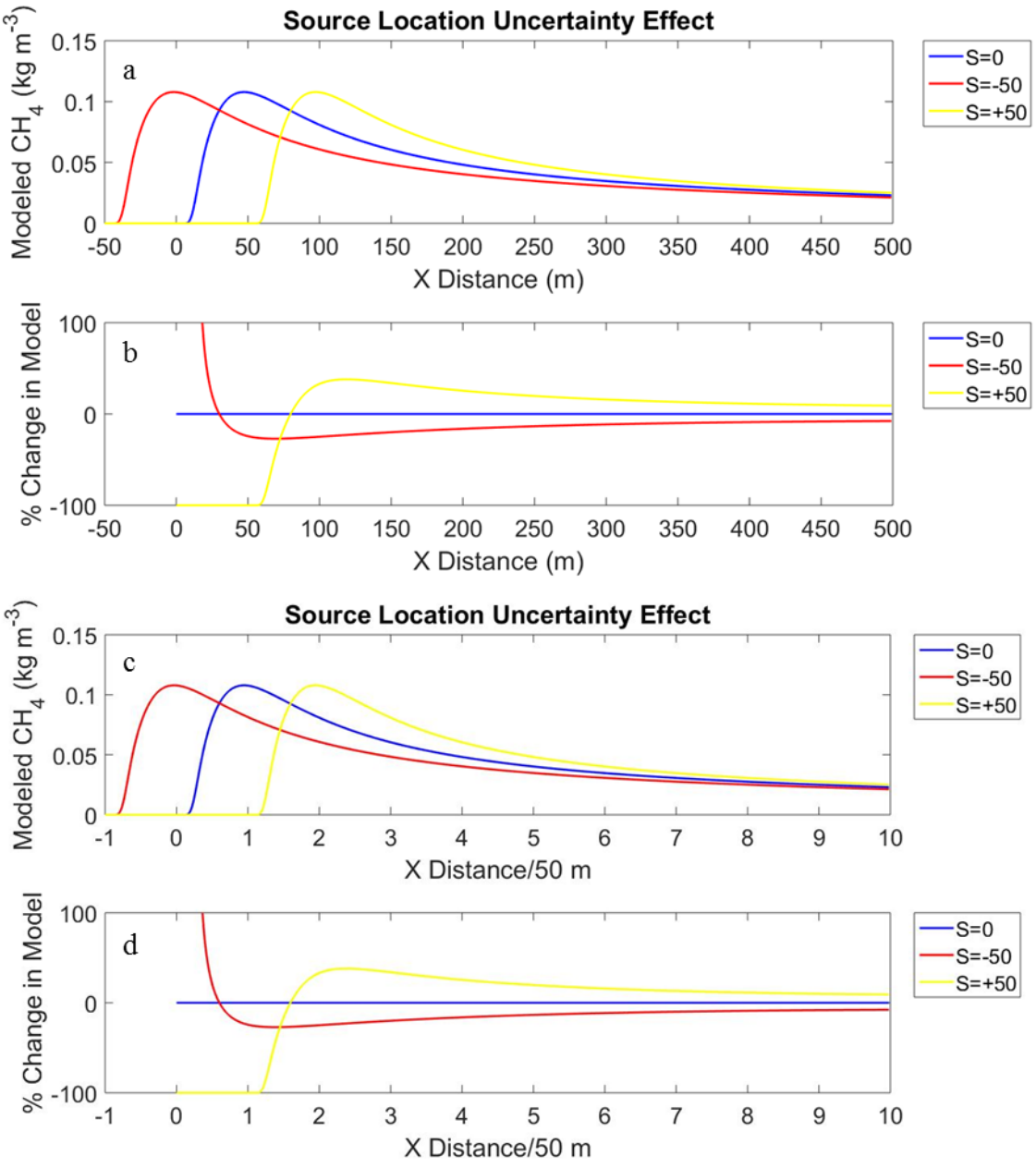
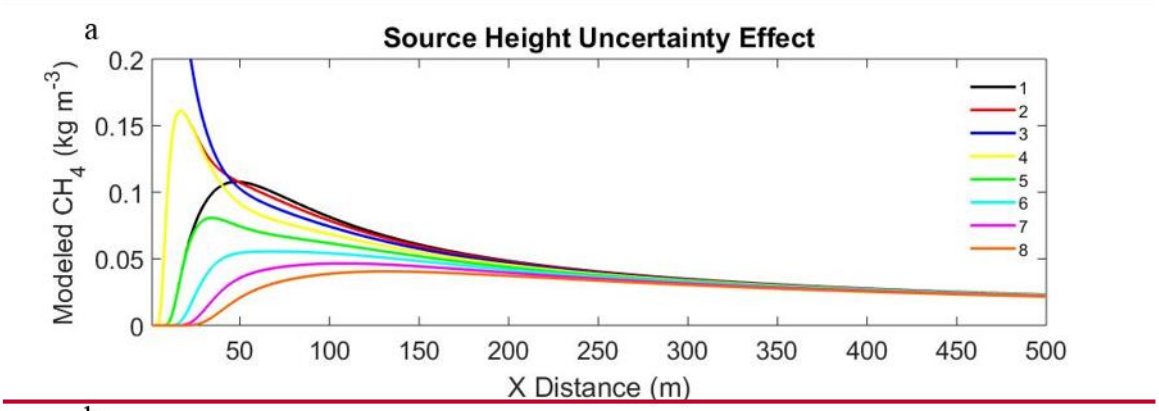
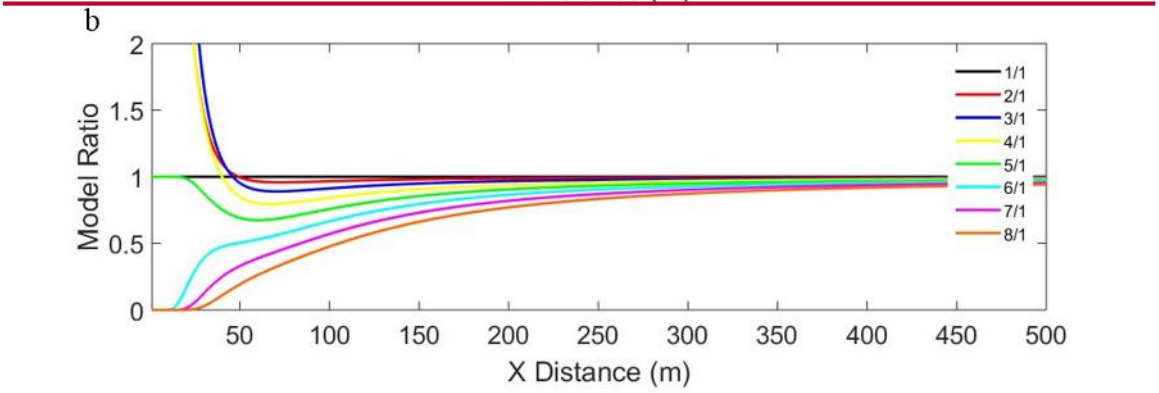


Figure ~~S10~~S12. (a) Modeled CH_4 at three source locations at different downwind x positions assuming 3 m receptor height, 1 m source height, neutral stability and 1.5 m s^{-1} wind speed. (b) ~~Ratio between the sum~~Change in ~~y and distance x~~% of the modeled CH_4 in each of the scenarios and 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.


a
Source Height Uncertainty Effect
Modeled CH4 (kg m-3)
0.2
0.15
0.1
0.05
0
50
100
150
200
250
300
350
400
450
500
X Distance (m)
1
2
3
4
5
6
7
8


b
Model Ratio
2
1.5
1
0.5
0
50
100
150
200
250
300
350
400
450
500
X Distance (m)
1/1
2/1
3/1
4/1
5/1
6/1
7/1
8/1

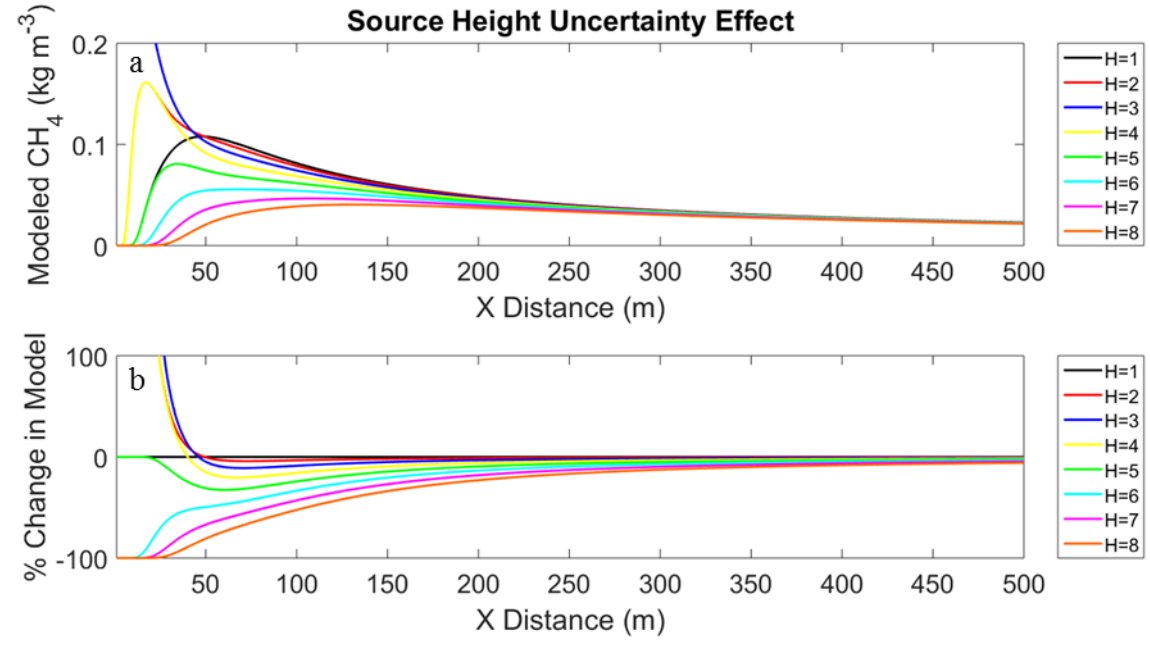
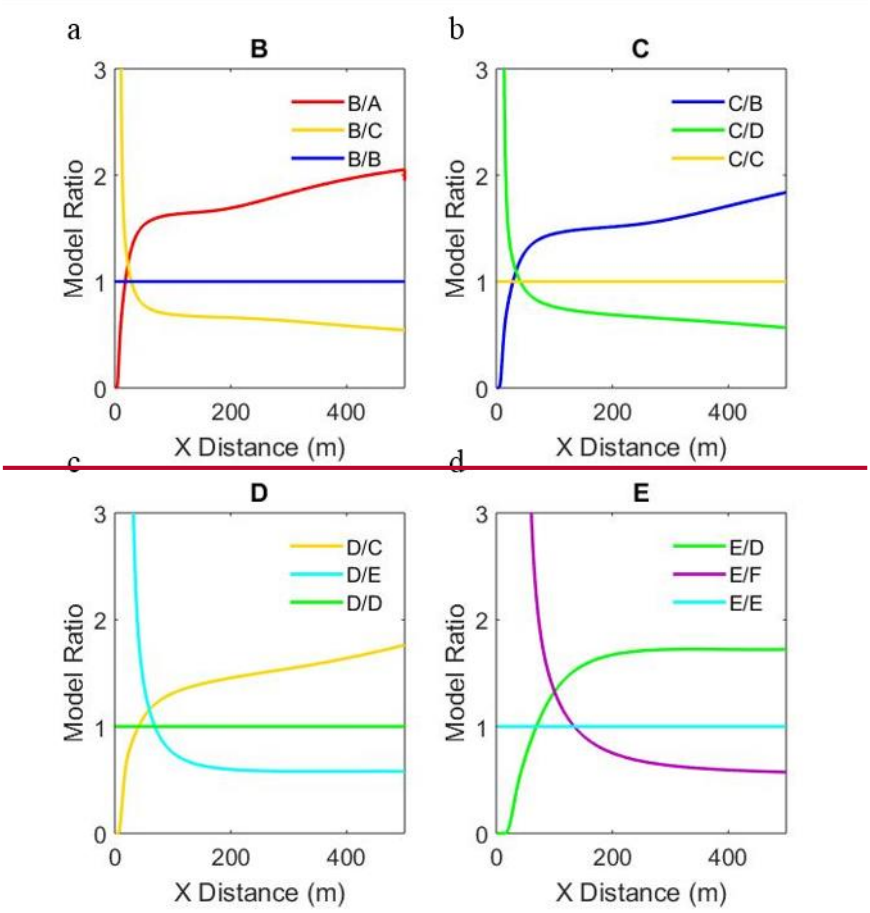

Source Height Uncertainty Effect
Modeled CH4 (kg m-3)
a
0.2
0.1
0
50
100
150
200
250
300
350
400
450
500
X Distance (m)
H=1
H=2
H=3
H=4
H=5
H=6
H=7
H=8
% Change in Model
b
100
0
-100
50
100
150
200
250
300
350
400
450
500
X Distance (m)
H=1
H=2
H=3
H=4
H=5
H=6
H=7
H=8

Figure ~~S11~~S13. (a) Modeled CH_4 using 8 source heights assuming 3 m receptor height, neutral stability, and 1.5 m s^{-1} wind speed. (b) ~~Ratio~~Change in % between the ~~sum~~modeled CH_4 for each of the scenarios and the base scenario (h=1 m).


a
B
B/A
B/C
B/B
Model Ratio
X Distance (m)
b
C
C/B
C/D
C/C
Model Ratio
X Distance (m)
c
D
D/C
D/E
D/D
Model Ratio
X Distance (m)
d
E
E/D
E/F
E/E
Model Ratio
X Distance (m)

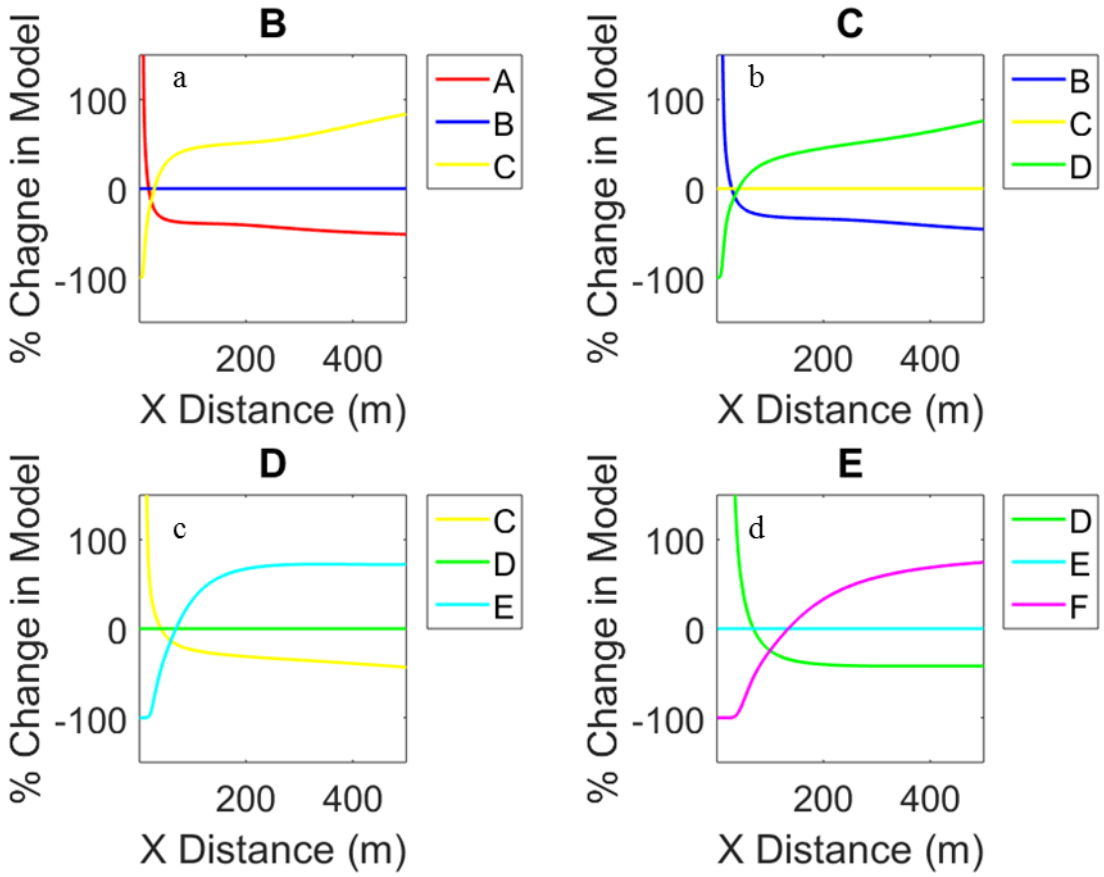


Figure ~~S12. Comparison~~S14. Change in percent between ~~Gaussian model sums and~~ modeled CH_4 between +/- one stability class and base scenario for (a) Class B, (b) Class C, (c) Class D, and (d) Class E assuming 3 m receptor height, 1 m source height, and 1.5 m s^{-1} wind speed.

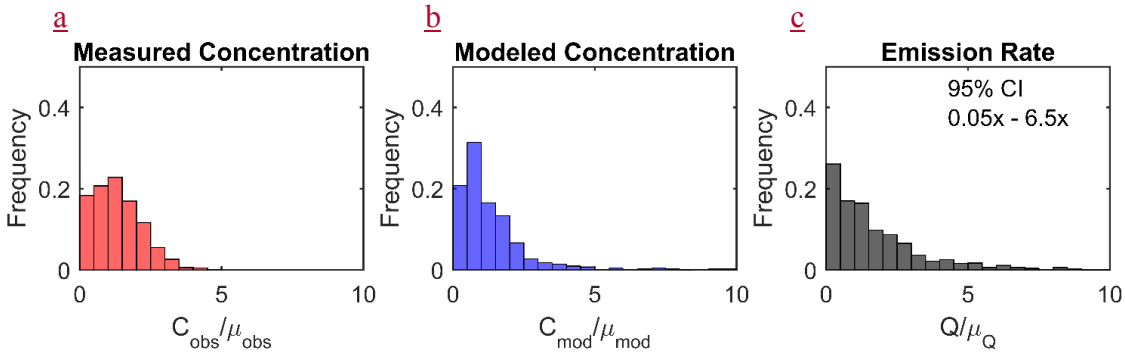


Figure S15. Monte Carlo simulations using 1,000 replicates for the standard sampling, 1 transect case showing (a) normalized observation distribution, (b) normalized model distribution and (c) normalized emission rate distribution.