

Egusphere-2025-5086
Response to Reviewer 2

Yamanouchi et al.

We thank the reviewer for the detailed comments. **We have updated the manuscript to address the issues raised.** This document summarizes our responses and documents the changes made to the manuscript. All reviewer comments are written in black font. All responses are written in **red font**, and all references to changes in the revised manuscript are written in **bold red font highlighted**.

General Comments:

1. Description of Experiments and Data Collection: More detail is needed to describe the data, including how the experiments were set up, what sensors were used, and how the sampling system was set up. Below is a list of parameters that absolutely need to be included. I suggest editing Table 1 to add some of these parameters.

- Release rate of controlled experiments

We agree that additional information on the controlled release experiments would be helpful. However, we wanted to keep the main text succinct and therefore added the comprehensive information in the appendix. As Table 1 contains information from both non-controlled release observations, we decided to not alter the table. The reported emissions from the controlled release data is given in the main text as well.

- What sensors (gps, ...)

Airmar mentioned in Section 2.1 also collects GPS coordinates. This was added in text.

- Sensor flow rates

The instrument flow rates is about 0.25 L/min. This was added in text.

- Distance downwind of the source which the transects were collected

This was added in the Appendix.

- # of transects

For the non-averaged runs (i.e., runs not discussed in Section 3.3), we are examining only 1 transect per inversion.

- Speed of mobile station

Speed ranges between about 50 to 70 km/h.

- Time sampling

Sampling frequency is 1Hz, and each transect takes about 3 to 5 minutes.

- Height of release

For the controlled release experiments, methane was released from the ground of the landfill.

2. Validity of HRRR Data: More description is needed that can help the reader follow and recreate what the authors perform. This work relies on outputs of this model to determine the dispersion parameters. In comparison to the time and spatial scales of the measurements, the resolution of HRRR is very coarse.

Thanks for raising this issue. The choice of HRRR was not based on it being the best choice, but its wide availability. We did however compare HRRR-based modelling results with modelling results using local observations. More details in response to part b.

a. Are the authors using output from HRRR (operational) or HRRR-x (experimental)?
Operational HRRR data was used, downloaded from NOAA cloud-based data servers (e.g., <https://registry.opendata.aws/noaa-hrrr-pds/>).

b. HRRR provides hourly output. Is that high enough resolution to meaningfully compare to in situ (1-Hz) measurements or transects that lasted for an unknown amount of time?
HRRR wind data showed performance that was on par or better than the on-board wind data from the weatherstation in the car; obviously, a dedicated, stationary anemometer set up inside of the plume would be ideal, but for our purposes, i.e. without site access and ability to deploy long-term weather monitoring sites, HRRR provides a product that is better than the onboard weatherstation (which need to be corrected for things like driving speed), while also providing the meteorological parameters needed for Obukhov length calculations. Hence, using HRRR allows to have consistent wind data and Obukhov length across most of Canada's densely populated regions.

c. Can HRRR provide valid wind data for a sampling height of 2.5 m agl? The lowest atmospheric layer in HRRR is typically ~6-12 m agl meaning that the lowest elevation wind parameter (e.g. U1000, V1000) is typically well above the sampling height presented. Broadly, it would be expected that wind speed at 2.5 m agl is less than that at 10 m implying that using HRRR will overestimate the emission rate. However, the impact is probably more complex than that and probably depends on the height of emissions and vertical velocity components, which are not currently discussed in this manuscript.

Again, we found that the wind data was similar to local data from airport weather stations or the mobile weatherstation data. It is also important to note that the

uncertainty of the Gaussian approach overall is significant and uncertainty in wind data is part of that.

d. Please provide the specific parameters used from the HRRR output.

The following parameters were used: Pressure, specific humidity, temperature, PBL height, friction velocity, ground heat flux.

e. General description of the parameters that the authors used from HRRR is needed. For example, Surface roughness in HRRR is assumed from the MYNN Surface Layer Scheme. In MYNN, that is largely dependent on land use, greenness fraction, and snow depth. This allows the reader/reviewer to assess whether these are reasonable parameterizations. Please include citations for HRRR

(<https://journals.ametsoc.org/view/journals/mwre/144/4/mwr-d-15-0242.1.xml>) and MYNN (<https://repository.library.noaa.gov/view/noaa/30605>)

Both citations and a list of HRRR parameters used were added in Section 2.2, and a note about surface roughness from HRRR assuming the MYNN scheme has been added in Section 3.1.

3. Gaussian Model and Fitting location: There are a number of assumptions and limitations within a Gaussian Plume model that are not addressed. Please see Seinfeld & Pandis "Atmospheric Chemistry and Physics" (chapter 18, see references therein). A very simplistic version of the Gaussian Plume equation that assumes total reflection at $z=0$ can be written as (S&P eq 18.88):

$$\Delta C(x, y, z) = \frac{Q}{U} \frac{1}{2\pi} \frac{\sigma_y(x) \sigma_z(x)}{\sigma_y(x)^2 + \frac{z^2}{Z(x)^2}} \exp\left[-\frac{y^2}{2\sigma_y(x)^2} - \frac{z^2}{2\sigma_z(x)^2} - \frac{z}{Z(x)}\right]$$

Where ΔC is the methane mixing ratio above background at location x, y, z . Q is the flux rate. U is the windspeed along flow of plume. $\sigma_y(x)$ and $\sigma_z(x)$ describe the Gaussian width horizontally across the plume and vertically at distance x from the source. $Z(x)$ is the height of the plume at distance x from the source. In this equation there are essentially 9 parameters: 4 that should be known (3 for the sample locations, 1 for the release height), 3 derived from HRRR (windspeed, $\sigma_y(x)$, $\sigma_z(x)$), 1 measured value ($\Delta C(x,y,z)$), and essentially 1 parameter (Q) that is solved for with the Bayesian approach. A similar equation must be used and all of the parameters need to be discussed in terms of what they are, how they influence the interpretation of Q , and how they are measured/modeled/parameterized.

a. What version of Gaussian Plume equation is used? Please provide the equation. Are you assuming 100% absorption by the ground or ground reflection? Reflection of the plume off the PBL? Please provide the Gaussian Plume equation.

The Gaussian plume model from the Polyphemus suite was used in this study. The complete model description is available in their paper (Mallet et al., 2007: 10.5194/acp-7-5479-2007). The model source code is available for download, citations also available in text.

b. How is vertical velocity accounted for? The authors mention “Briggs plume-rise scheme” (L144), but do not provide a citation. How is a reviewer or reader supposed to reproduce this?

Reference to Turner (1969) was added. The Polyphemus model allows the operator to choose to use the Briggs plume-rise scheme. We have made all of our data and modeling/inversion/analysis code available (cited with DOI in the paper), to ensure that others can easily reproduce our results. Model code is available at:

c. Specify, maybe in a table, the values of parameters that are modeled/derived/parameterized, including $s_y(x)$, $s_z(x)$, plume height at the transect ($Z(x)$) for each transect measurement set? What are their uncertainties? How do these uncertainties affect the interpretation of Q?

Thanks for the suggestion, we now provide more information. A table for calculating the dispersion coefficients has been **added to the Appendix**. In the Polyphemus framework used in this study, parameters such as (S_x) and related Gaussian plume dispersion coefficients are not explicitly stored as output variables. Instead, the model outputs concentrations of the transported species (in this study methane) which are then used for subsequent analysis.

The primary objective of this study is to improve the quantification of methane emission rates rather than to perform a detailed tuning of the Gaussian plume parameterization itself. As such, our focus is on the resulting concentration fields and their relationship to inferred emissions, rather than on intermediate dispersion parameters.

Regarding vertical resolution, the Polyphemus model allows for flexible specification of vertical layers. In this work, we selected the model layer corresponding to the sampling height of our measurements. This choice was made following sensitivity tests in which different vertical layer configurations were evaluated, and the configuration providing the best agreement with observations was retained. **This discussion was added to the paper to improve clarity.**

d. Why initiate the model for the selected parameters, e.g. 1 g (CH₄)/sec? How would changing to a different initialization state affect the results? My guess is that because the Bayesian approach solves for only 1 parameter, it is insensitive to the initialization state.

This prior was chosen as it was roughly the typical value of what we have seen from other facilities we have observed. We could have chosen “better” a priori values by adjusting them for each site if emission rates had been known. However, both for the controlled release as well as the industrial sites surveys true emissions were not known to us during the experiments, hence, we used the same value for all inversions for consistency and to not color our results. Prior to the study discussed in this paper, we did perform an analysis by changing the a priori, and found that it showed little to no differences in the estimated emission rates.

e. Practically, it was unlikely that the transects were perfectly orthogonal to the direction of flow (wind direction). How is the non-orthogonal transects accounted for? If plumes were perfectly Gaussian, then a non-orthogonal transect would introduce skewness to the measured concentration profile.

This is accounted for in our modeling and inversion methodology; because we model the plume as a grid, if the transect is not orthogonal, the modeled transect will “cut” through the plume at an angle, and will indeed have skewness. In other words, our forward model can provide forecasts of non-Gaussian plume shapes on non-orthogonal transects.

f. How was weighting b/w stability classes performed? i. Which set of stability classes are used? I can infer from relation to surface roughness and Obukhov length, but there are other classifications that are in use, including; US EPA 1970 (“Workbook of atmospheric dispersion estimates”), Gifford 1961 (“Use of Routine meteorological observations for estimating atmospheric dispersion”), etc. ii. Weighting is performed b/w 2 parameters with different operational ranges ($z_0 \sim 0-0.6$ versus $1/L \sim -.15 - 0.1$). Is one value weighted more than the other? Is it simple Euclidean distance between stability classes? What is being weighted, the dispersion coefficients or the resulting Q of each class?

Thank you for raising this issue. We have clarified the issue in the manuscript. The weights are calculated by the distances to the nearest stability curves. The distances were calculated holding z_0 (y-axis) constant (i.e., the horizontal distance), as surface roughness is assumed not to change (at least not in the time span relevant to these analyses).

g. How do the authors account for the averaging time of their measurements? Most stability classifications are based on average distribution of the emission across multiple hours (see Briggs 1972, Pasquill 1971, etc). This will result in a much broader plume distribution than would be observed if one could instantaneously sample the complete transect. A quote from Briggs 1972 exemplifies this: “However, the comparison made

with the range of β values observed by Hoult et al (1969) is not a fair one since these values represent best fits to many individual data, while the Schwartz and Tulin β values are best fits to several sets of averaged data.” Below is a figure from Seinfeld and Pandis demonstrates the increasing plume width with increasing measurement time.

Figure 1: Seinfeld & Pandis, "Atmospheric Chemistry and Physics" Using dispersion constants that are derived from long sampling periods will overestimate the dispersion and width of the Gaussian shape used to fit the data. The result will be an overestimate of emissions as the model things those emissions are spread over a larger area.

Please see next comment.

h. How do the authors account for the non-instantaneous measurement of the transect? As the vehicle proceeds along the transect, the wind is also changing direction meaning that the mobile platform is never sampling the same plume distribution. This is a major hurdle for all mobile platforms and needs to be discussed. Long sampling periods with many transects will be essential, but the sampling time and number of transects are never discussed.

We thank the reviewer for raising this important point regarding plume meandering and dispersion, as discussed in classical formulations such as those of Briggs and Pasquill.

In principle, these effects, particularly wind meandering and enhanced lateral spread, are expected to influence plume behavior and could lead to biases in Gaussian plume-based inversions. As noted in previous work (e.g., Ars et al., 2017), and confirmed by our own analyses, the Gaussian plume inversion methodology tends to *underestimate* emissions when compared against observations.

This empirical finding suggests that, despite the theoretical expectation that additional dispersion processes (e.g., meandering) might broaden the plume and potentially increase inferred emissions, the net effect in practical applications of the method is an underestimation bias.

We also note that we, although not discussed in the paper, also tried using OTM-33A. This technique leverages plume meandering by using longer duration and stationary observations in close proximity to the site to estimate emissions. For our mobile surveys we have chosen greater distances where these effects appear to play a much smaller role; indeed, we typically use OTM-33A in cases where no roads downwind of the site are available (only right by the site), and Gaussian model inversion if roads farther away are available to collect transects. When choosing transects too close to the site, plume shapes often appear non-gaussian, but for quantification purposes we choose an appropriate distance so the observed transects are mostly Gaussian. Nevertheless, we recognize that real world measurements are never really truly Gaussian.

The primary objective of this study is to improve the estimation of methane emission rates in an easy to use and fairly widely applicable approach rather than to fully resolve or parameterize all atmospheric dispersion processes. In this context, our methodology is designed as an improvement upon established Gaussian plume inversion approaches (e.g., Ars et al., 2017), and we evaluate its performance empirically against observations.

We have clarified this point in the revised manuscript, explicitly mentioning that averaging transects may help mitigate biases due to the non-instantaneousness of measurements.

4. Results & Discussion: This section should be restructured and address critiques provided in General Comment 3.

a. Presenting more of the transects in figures would improve communication of the technique and provide more transparency (see comment about Figure 2). While the authors show data from transects of 3 successful uses of their approach (Fig 2, 3, 5), they do not show any data from the other transects. There are 14 datasets and several are discussed (1242-257) that did not behave as those simple, easy to interpret datasets.

Transect figures from all 14 runs are now in the appendix.

b. Discussion of fitting of source location is not particularly impactful and, in my opinion, does not justify the amount of space that is discussed here. It is intuitive that the plume axis (center) is observed at a different location than the wind data implies. Significant uncertainty in wind direction arises from using the HRRR datasets: 1) Winds are rarely uniformly linear as implied by using 1 hr HRRR wind data, and 2) the 3km average likely does not represent ground-level winds that are influenced by local features like buildings, trees, etc.

The discussion on source location translation is included for completeness, as it represents a potential source of improvement in Gaussian plume inversions, especially regarding previous findings that indicate that inversions nearly always underestimate emission rates.

We evaluated multiple sources of wind data and found that the HRRR dataset provides wind estimates that are as good as, or in some cases better than, those obtained from our onboard (mobile) meteorological measurements. While it is true that a dedicated, well-sited stationary anemometer would certainly provide higher-quality wind observations, such instrumentation is not always feasible in the context of our measurement campaigns.

Our work is motivated by practical applications such as surveys of oil and gas facilities, landfills, and similar sites, where rapid deployment and spatial coverage are required. In

these scenarios, it is not always possible to deploy more sophisticated, stationary measurement setups. We therefore rely on wind fields that are both readily available and robust in practice. Based on our analysis, the HRRR dataset satisfies these requirements and performs at least as well as the onboard system used in this study.

c. Recommend restructuring the results with subsections for 1) Controlled releases, 2) several possible sections discussing limitations of approach, 3) application to facility data.

We thank the reviewer for this suggestion; as the focus of this paper is the methodologies (of source translation and stability class blending), and the observational data is an illustrative demonstration of the application of said methodologies, we feel that our current structure is better suited (for our focus and goals of this paper).

With that said, we are currently preparing a follow-up paper which *does* focus on observational data (with the strategies discussed in this paper applied), and we will structure that paper with this recommendation in mind.

Specific Comments:

Pg 1 L 38-57 This paragraph discusses how global methane budget is estimated using top-down or bottom-up approaches. This seems tangential to this work and doesn't promote the development of Gaussian Plume Models. Suggest removing.

The intent of this paragraph is to provide essential context and motivation for our work. Our study is ultimately aimed at improving methane modeling in support of emission quantification, which is a central objective in atmospheric methane research, particularly for our (governmental environmental research agency) research. As such, briefly outlining the range of existing methane emission quantification approaches (e.g., bottom-up inventories and top-down atmospheric methods) helps situate our modeling framework within the broader scientific and regulatory context in which it is applied.

Pg 3 L 65-66 "...stead-state solution with constant emission rates as well as constant and homogenous winds and terrain." Technically, the model provides a multi-hour averaged dispersion. It is not really a "steady state" and doesn't assume "homogenous winds". See references in General Comment 3g.

The wording has been updated to to clarify this.

L 87 None of the listed sites in section 2.3 are described as landfills?

Petrolia site is a landfill. **The wording in the manuscript was changed to clarify this.**

L 87-89 Please help describe the measurement process in more clarity. Are some of these sample sets only a single transect? See general comment 1.

The transects discussed outside of Section 3.3 (Averaged Plumes) are singular transects.

Pg 4 L 96 The sampling system is on top of a car and only 2.5 m agl? That seems really low and only ~1 m above the car. At this height, wouldn't the car affect the windflow and measured concentration? How tall is the vegetation and other structures in the vicinity? 2.5 m agl is correct. The car is a passenger vehicle, and the weather station is mounted on a pole about 1 m above the vehicle's roof on the front. Similar setups are frequently used and previous studies found that this type platform is able to detect and quantify methane in various environments, including during controlled released experiments (see e.g., Tettenborn et al., 2024: <https://doi.org/10.5194/amt-18-3569-2025>) in rural areas, e.g., near landfills (Gillespie et al., 2025: <https://doi.org/10.1021/acsestair.5c00127>), and in urban areas (Vogel et al., 2024: <https://doi.org/10.1021/acs.est.3c03160>).

Pg 5 L 125 "... stability class, accounting for 48% of the uncertainty, followed by wind direction (38%), windspeed (28%), and source location (19%)." The sum of these adds up to >100% of uncertainty. How is that possible? Do the authors mean that there is a 48% uncertainty in the source estimate due to stability class?

These uncertainties are based on calculation by Ars et al. (2017) for the same model. So the 48% indeed refers to the uncertainty component in emission rates from stability class choice. These numbers were taken from Ars et al. (2017). As these errors are orthogonal, they should be added in quadrature, leading to their overall uncertainty of 75%.

L135- 136 Please provide complete list of model variables used and be specific to what they are called in the datafiles.

The variables used, and their respective name (both the "shortName" and "fullName") in the HRRR datafile are as follows (shortName in parenthesis):

Atmospheric pressure: Pressure (PRES)

Specific humidity: Specific humidity (SPFH)

Temperature: Temperature (TMP)

PBL height: Planetary Boundary Layer Height (HPBL)

Frictional velocity: Frictional Velocity (FRICV)

Ground heat flux: Ground Heat Flux (GFLUX)

z0 (surface roughness): Surface Roughness (SFCR)

Wind speed (u): U-Component of Wind (UGRD)

Wind speed (v): V-Component of Wind (VGRD)

L144 Technically, stability class does not affect plume-rise. Plume rise is affected by wind fields, buoyancy (e.g. composition and temperature of plume), solar heating and

albedo, etc. Stability class is a qualitative description of things like plume rise and turbulence, etc.

We agree that atmospheric stability class is not itself a direct physical driver, but rather a classification that reflects underlying meteorological conditions.

Our intent in this sentence was to convey that the stability class implicitly captures information about factors such as solar heating and turbulent mixing, which do influence plume rise. As the reviewer correctly notes, solar heating (flux) for example plays a key role in plume dynamics, and this is one of the parameters used in determining stability class(es).

L144 "...Briggs plume-rise scheme..." Please provide a reference!

L146 Define which set of stability classes you use. Provide a reference!

Reference to Turner (1969) was added. This is the same article cited by Mallet et al. (2007) (the publication for the Polyphemus model development).

L153 Surface roughness and Obukhov length were not listed as variables from HRRR earlier? Please explain where these parameters come from and how they are derived (see General Comment 2). This allows the reader to assess validity and limitations of using the HRRR data without digging through other literature. Oh, and provide a reference to the HRRR model and MYNN model used by HRRRv4.

The stability class(es) were derived by using surface roughness from HRRR, and calculating L (or rather 1/L) using meteorological parameters in HRRR, by calculating virtual temperature (and potential temperature), kinematic energy flux, convective velocity scale, and using PBLH from HRRR (please see above comment for the HRRR parameters as well as <https://test.pypi.org/project/blendie/>).

P6 L154 How was the weighting performed? See General Comment 3f.

Please see above comments.

L165 "...translated along (parallel to) the transect path..." both of these descriptions are a little confusing.

This simply means that the linear translation was done in the direction parallel to the transect. **The word "linearly" was added (before the "translated along (parallel to)..." line) in the paper to clarify.**

L166 This is not what autocorrelation describes. Autocorrelation describes how a signal or time-series correlates with a delayed version of itself (i.e. 'auto'). Not the phasing between two separate signals (observed and modeled concentration).

Thanks for catching this. A better wording would be cross-correlation. Specifically, we used the `scipy.signal.correlate()` function

(<https://docs.scipy.org/doc/scipy/reference/generated/scipy.signal.correlate.html>) was used. The elements in the output array from this function show how similar the two sequences are as one is moved relative to the other. **This wording was corrected in the paper.**

L167 How is skewness incorporated into a Gaussian Plume Model? The angle of the transect with relation to the wind direction is not discussed anywhere in this manuscript. Since the model output is a grid, if the transect path is “diagonal”, the transect will naturally capture the skewness. **This is now explicitly discussed in the paper.**

P7 L170 “Autocorrelation values of 0.66 to 0.955 were chosen as cutoffs...” Again, incorrect use of ‘Autocorrelation’. Assuming that the authors mean ‘correlation’, how were a range of values chosen as a cutoff? Did they test different cutoffs affected their interpretation of Q? Maybe using the controlled release? That would be good to do and discuss. Wouldn’t a simple Least-Squares approach achieve the same thing without tuning a cutoff? L187- 189 Were the other datasets only single transects? Describe how these transects were average together. “Overlaying the measurements on top of each other” is not a precise description.

This wording was corrected in the paper.

Sect 2.3 Really need to know parameters like the distance of transects downwind of source, number of transects for each experiment/measurement set, speed of the mobile station, amount of time sampling the plume. These are essential parameters in the Gaussian Plume Model (see General Comment 3).

Table with the distances from the sources have been **added in the Appendix.** Transects discussed outside of Section 3.3 (Averaged Plumes) are singular transects. Speed for the transects were about 30-50 km/h.

P8 L205 Correlation (R2) is a poor metric for determining performance. Hypothetically, assuming two Gaussian shapes with the same mean but different standard deviations and amplitudes that represent our measured and model transects. These two profiles will always have a high R2 value because they always increase and decrease together. They may not even have a slope near 1 since correlation doesn’t care about that. Maybe Chi2 metric better describes the predictiveness of the model.

We acknowledge that R2 is not a perfect indicator of performance. However, we have found that examining R2 is a useful tool in quality control of modeling/emission estimation.

The figure below (Figure 1) compares R2 and Chi2 from the smoothed, averaged controlled release experiments (data seen in Figure 6 on the paper), ordered from the runs with the highest R2 to the lowest. To illustrate this point, Figure 2 below are the

runs from 0517_10 and 0517_5, respectively. 0517_10 has the second highest R2, while 0517_5 has the lowest, and it is visually clear that 0517_10 has a better fit; R2 was able to discern this, while Chi2 was not.

We fully agree with the reviewer that *in theory*, two Gaussians with the same mean (but different sigma and amplitude) will always have high R2. However, experimental observations tell a different story; we have found that R2 is able to give us useful insight into how well the model fits the data, even in situations where the only difference in the model is the spread and amplitude. This is evident in the figure below (Figure 3); the plots show two modeled runs of the same transect, but ran with stability classes A and E (hence same peak, but different widths and amplitudes). The R2 for A was 0.51, but for E, it was 0.88, a noticeable increase.

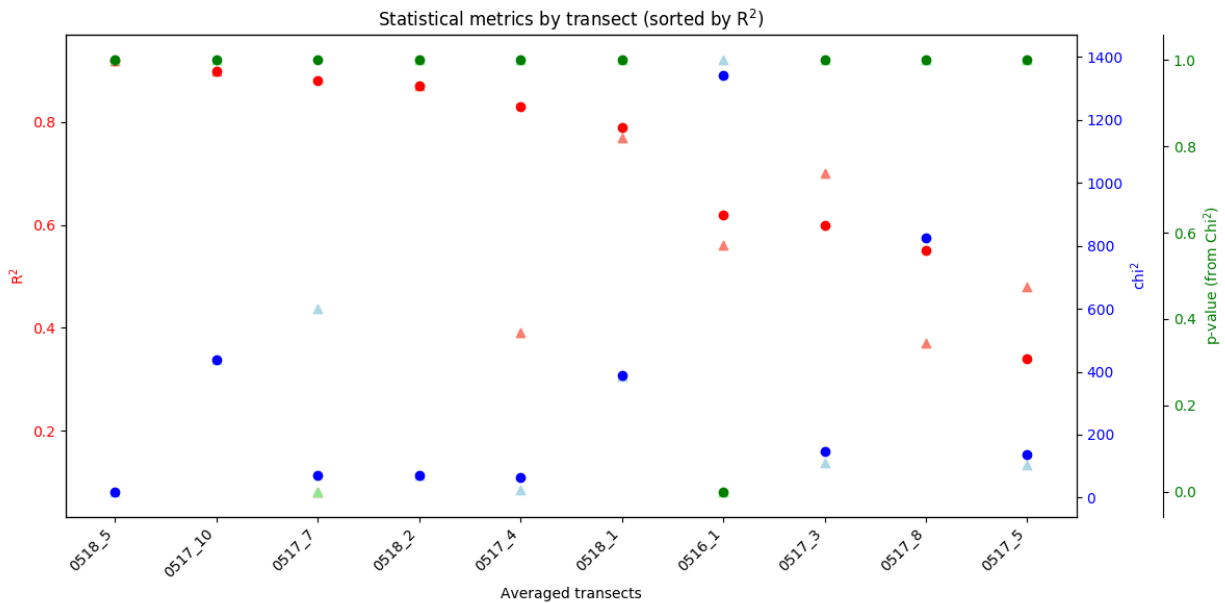


Figure 1. Statistical metrics of averaged transects (averaged controlled release data). Optimized location datapoints are indicated by the circular dots, and the unoptimized (original) location datapoints are indicated by the triangular points. Red/pink points indicate R² data, blue the chi², and green the p-values.

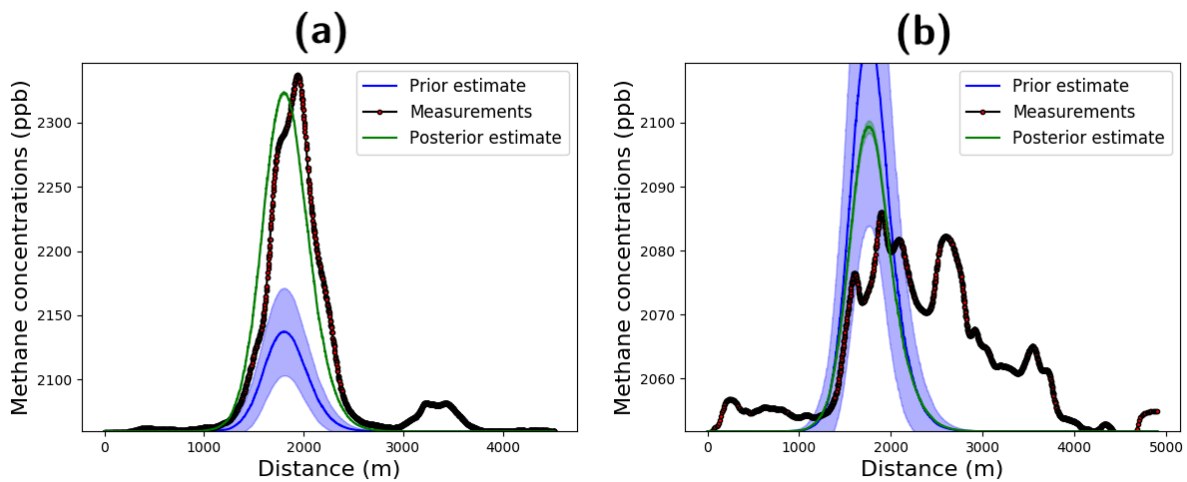


Figure 2. Transect plots for 0517_10 (a) and 0517_5 (b).

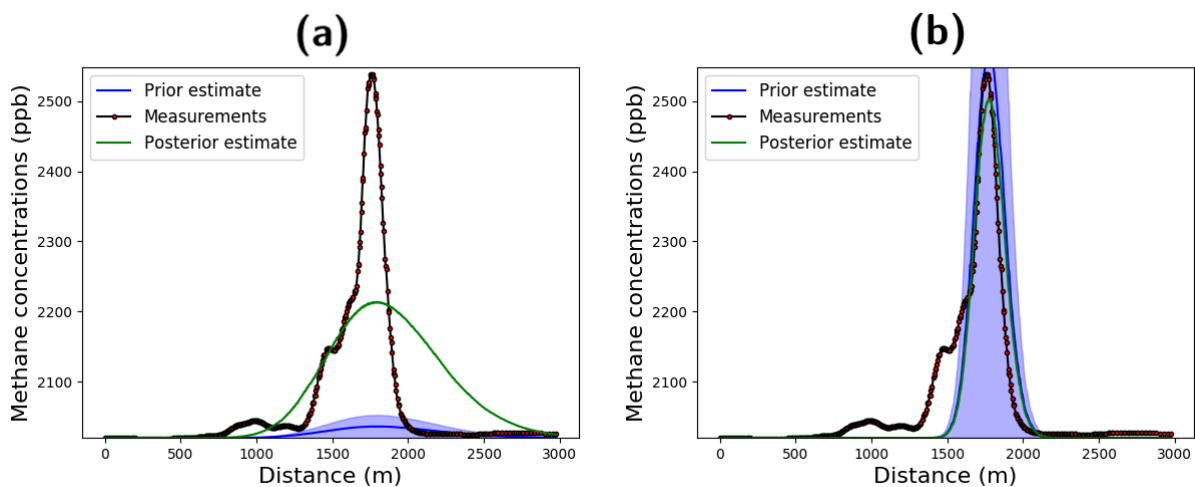


Figure 3. Same transect, but modeled with stability class A (a) and E (b).

L211 "...converged to give physical and meaningful emission estimates..." This does not really provide meaning without defining what a physical and a meaningful emission is. Is it within a certain % of the true emission rate?

We believe that this plume, while it was observed during one of our controlled release experiments, was not from a controlled release experiment, as the source adjusting algorithm located the source to be over a kilometer away. Thus, we cannot compare to a "true" emission rate. We say that the result was meaningful because they gave us an estimate on emission rate, and we included in the discussion that it was "physical"

because it was reasonable (the estimated rate is inline with emission rates typically seen from farms or waste treatment facilities).

L222 "...extremely sensitive to..." This doesn't seem true. It is insensitive to z_0 , but is also insensitive to Obukhov length in areas with $z_0 > 0.2$, (e.g. $z_0 = 0.4$, $1/L = -0.10$). The sensitivity is only apparent for the neutral and near-neutral stability classes. **This was corrected in the paper.**

Sect 3.2 See general comment 3 and 4.
Please see above comments.

L226 "The source translation algorithm generally resulted in higher correlation, lower RMSE, and higher emission estimates." By definition, aligning the peak in the measured concentration with the model plume center will maximize the estimate of the Gaussian Amplitude and emission estimate. This should not need discussion.

We agree that this interpretation may be apparent to experienced readers. The intent of this sentence is not to introduce a novel finding, but rather to introduce (explicitly) the main takeaway of this section (hence it is the first sentence of the section) and to guide the reader in understanding the effect of the source translation.

L228 What was the angular offset between the HRRR wind direction and the measured plume axis?

This information has been added to the Appendix, along with the downwind distances.

P9 Fig 2 The 3 panels should be combined into a single panel with the different posterior estimates having different line styles. This would reduce redundancy of the figure and allow better comparison of the blended dispersion estimate to the stability classes. It would also be good to see the uncertainty bands of the measured data. Provide information on interpreted emission rate from each posterior and the known emission rate of the controlled release experiment.

Figure 2 has been updated to show the model posterior from the 2 stability classes and the blended run in one figure. Measurement uncertainty and emission rates for the controlled release experiments are both discussed in text.

L238- 239 "This is supported by the fact that the overestimation was present before the source translation took place." I'm confused by why this supports your hypothesis of uncertainty in downwind distance from the source.

We overestimated the emission rate regardless of the source location translation (i.e., including with the "non optimized" initial location). If the modeled source location is

farther than the true source location, then the model will need to increase the emission rate to match the observation. We are suggesting that this may explain, at least in part, the overestimation, particularly given that Gaussian dispersion model inversion tends to underestimate (as discussed in the paper) emission rates.

L242 What is meant by “real” observation data?

This simply means not from controlled release experiments, but rather data collected by driving through plumes emitted from functioning industrial complexes, landfills etc.

L242- 258 It would be good to show some of these transects, similar to what is suggested for Fig 2. Maybe with a subplot for each of the non-ideal datasets discussed here.

Please note that all transect figures are **now available in the appendix.**

P10 Fig 3 See General Comment 4. This figure is probably not needed. Showing the transect before aligning the model to the data is not very interesting.

While the results of translating the source may be obvious to experienced readers, we included this figure to explicitly and more importantly *visually* illustrate the alignment of the model to the observation.

L250- 251 “Note that...thus emissions are not expected to agree with each other.” I am really confused by this statement. Maybe better description of the data collection would help.

The site referred to as Courtright in this study is a very large industrial complex, with likely multiple sources within it. In contrast with controlled releases where each release experiment should (in theory) have the same constant release rate, these real world sources, especially large scale facilities, will likely have multiple sources, each with different emission rates. **Manuscript was updated to clarify this.**

L255 “...size of the signal...” -> “emission rate”

Fixed.

L257 “...returned meaningful and physical results” What is a meaningful result? What is a physical result?

The results were meaningful because they gave us estimates on emission rates, and we included in the discussion that they were physical because they were reasonable (that is to say, given the size of the facility, emission rates were reasonable).

L260 "...plume shape is the same whether the source is farther away and emitting less, or closer but emitting more..." This isn't true. The source that is farther away would have a much smaller amplitude and wider distribution than the nearer+larger source. Thanks for catching this, we had the sentence backwards. **This mistake has been corrected in the paper.**

L261- 270 This section doesn't make much sense to me

Please see the figure below (Figure 4). The figure depicts obtaining 2 transects from the same source. This allows for constraining the source in 2 dimensions (i.e., x and y). By iterating this process, we are suggesting that a better estimate on source location may be obtained. The solid cyan and pink arrows by the star show the translation vectors for the first and 2nd transects, and if this process is repeated, we are suggesting that a spiraling (and narrowing) stairwell-like shape that zeroes in on the "true" source location should emerge.

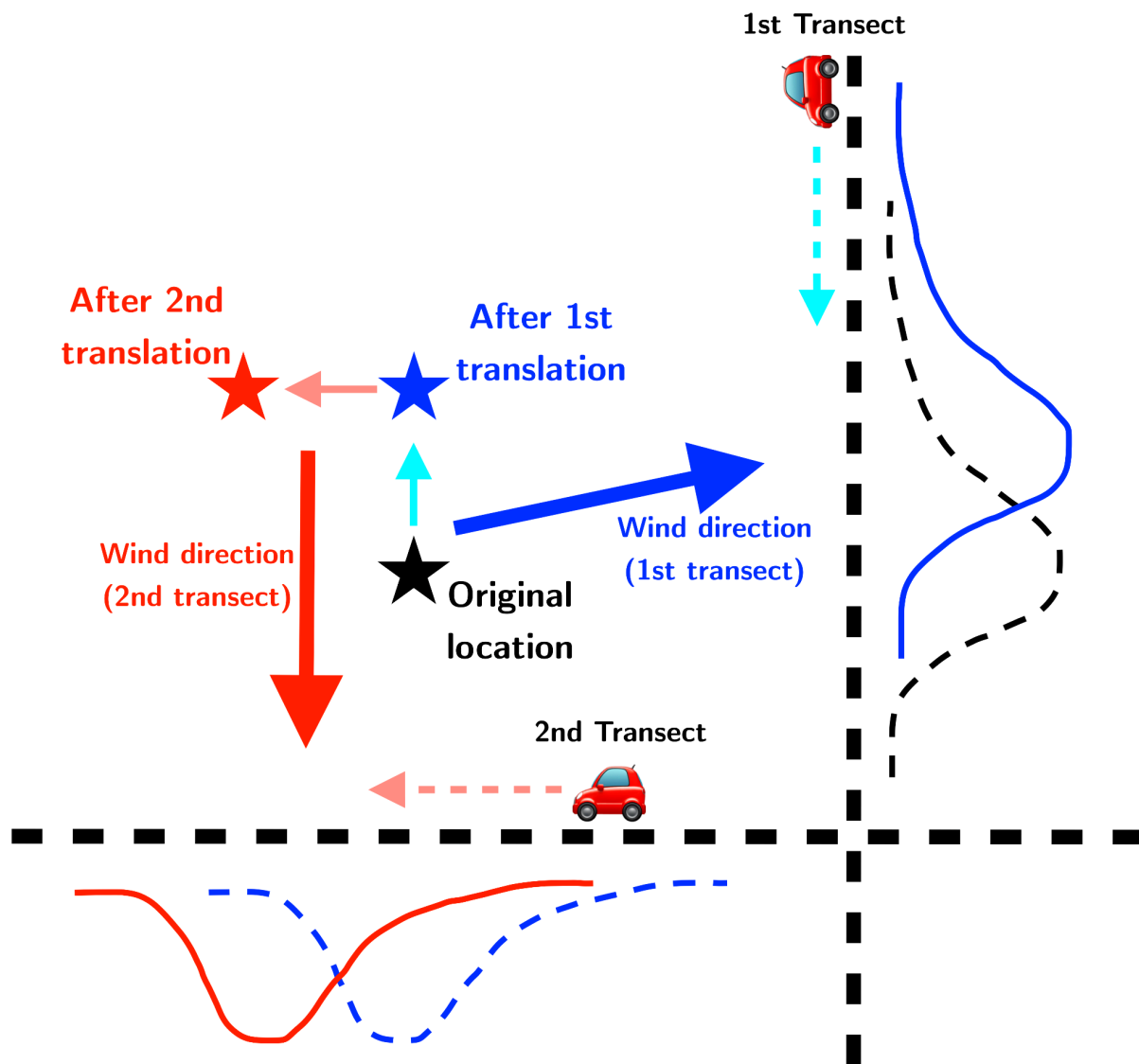


Figure 4. Schematic for source translation.

P11 Sect 3.3 See General comment 3 regarding derivation of dispersion parameters. Table for calculating the dispersion coefficients has been **added to the Appendix**.

P12 Fig 4 Add a line showing the transects. Add heat map for measured CH4 concentration Move legend over map in an area that doesn't conflict with location and concentration data. Mark a priori wind direction and plume locations.

The figure has been updated to show the transects (observations), and the legend was moved to the corner to show the transects better.

L276 What is the time delay for air to move through the tube? What size tube? How long? It can't be that long but that may depend on speed of the mobile station.

It takes 12 seconds for the air to move through the tube, this delay is corrected for during data post processing. The sampling line is composed of about 2.5 m of 1/4-inch tubing.

L281 Why remove this point? It is hard to classify this as an outlier since it still falls within the 95% CI of the regression.

This point was dropped as its emission rate was an order of magnitude higher than other data points. Furthermore, when we examined the transect and model, the plume transect lacked a Gaussian shape and it was clearly not a good data point. Had this been an observation that we took outside of a controlled release experiment, it would not have met our quality control standards and would have been dropped. We kept this in the analysis and mentioned it here as a point of discussion. However, it was included in the figure, and as discussed in text, correlation was not severely affected with or without the inclusion of this point.

P13 Fig 5 • Add axis labels for subplots c and d. • Use a constant font size. The subplot labels are much larger than other fonts. The font used for the axis ticks and colorbar are too small. • It doesn't really seem hard to fit a second Gaussian curve to the residual of the primary Gaussian. This would allow identification (and quantification) of multiple peaks observed along the transect.

Figure 5 has been updated.

We thank the reviewer for this suggestion. We are currently working on doing inversions with multiple sources to identify and quantify multiple sources simultaneously. However, as the primary focus of this paper is to showcase the source (singular) translating algorithm and the stability class blending methodology, we have elected not to discuss multiple source quantification.

L289- 291 "This is likely because the averaging..." Again, this is the expected outcome of averaging transects and shifting model peak location.

We agree that this interpretation may be apparent to experienced readers. The intent of this sentence is not to introduce a novel finding, but rather to explicitly clarify our interpretation of the results and to guide the reader in understanding the effect of averaging on the observed behavior.

P14 Table 2 • R2 is not a good performance metric. • Add columns for known emission rates of experiments for controlled release experiments. • The "before" columns are not particularly interesting. See General Comment 4.

Please see above comment regarding our use of R2. Also, as Table 2 contains information from non-controlled release observations, we have not added this in the Table itself.

P15 Fig 6 Why do the ranges of both axis go below 0 kg/hr?

The axis ranges in Fig. 6 were chosen to improve the readability of the plot. In particular, placing the zero value directly at the axis boundary resulted in the zero tick label being partially obscured, which made it more difficult to identify and interpret the reference point.

To enhance clarity, we extended the axis range slightly beyond zero so that the zero tick is fully visible. This adjustment is purely visual and does not affect the data or its interpretation.

P16 L326- 331 This still doesn't make sense to me.

Please see above comment and figure (Figure 4).

