General Comments:

Overall, we think this is a rigorous methods paper and does a good job of describing a novel combination of tracer and dispersion methodologies. The descriptions of the experiments and methods are clear, and will be well understood by the community, both on the experimental side and on the dispersion side. We think this paper should be published after addressing the comments below.

We thank S. C. Herndon, J. R. Roscioli and T. I. Yacovitch for this general assessment of our study and manuscript and for their constructive and technical comments that will help improve both our analysis and the text of our manuscript.

1. Please cite some important related research that explores tracer mislocation effects:

Goetz et al. used a similar combined methodology in 2015 for tracer release experiments at wellpads. Goetz, J. D.; Floerchinger, C.; Fortner, E. C.; Wormhoudt, J.; Mas- soli, P.; Knighton, W. B.; Herndon, S. C.; Kolb, C. E.; Knipping, E.; Shaw, S. L.; et al. Atmospheric Emission Characterization of Marcellus Shale Natural Gas Development Sites. Environ. Sci. Technol. 2015, 49, (11), 7012; DOI: 10.1021/acs.est.5b00452.

This article will be cited in the section 2.5 to mention the possibility of using a Gaussian model to determinate a correction factor in order to take into account the mislocation of the tracer in the tracer released method.

Roscioli et al. Performed an extensive error analysis of the impact of tracer mislocation using dual-tracer release methodology. Additional methods of calculating the methane/tracer ratio are also described. 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.; et al. Measurements of methane emissions from natural gas gathering facilities and processing plants: measurement methods. Atmos. Meas. Tech. 2015, 8, (5), 2017; DOI: 10.5194/amt-8-2017-2015.

We will cite this article in section 2.2 to present the different methods of calculating the methane/tracer ratio and the impact of the tracer mislocation in the dual-tracer estimates. This article and the one suggested above will also likely be cited in the new conclusion too since the latter will be expanded into a sort of discussion / conclusion section (in order to include some of the discussions that previously lengthened sections 2-4 of the present version of the manuscript, and in order to lead the discussions asked for by the four reviewers).

2. Please address some potential problems with the experimental measurements: Have experimentally measured mixing ratios been calibrated? Please briefly mention calibration procedure.

Before the experiment, the instrument has been tested in the laboratory. It showed a good linearity over a large range of mixing ratios and a good stability over time with small dependency to pressure and temperature. The feedback from other instruments of the same type was that after a shutdown an offset could appear. To control for that offset, we measured a gas with a known mixing ratio (calibrated with a multi-point calibration in the laboratory) before each series of measurements in order to ensure the good analytical performances of our instrument. No offset was seen after shutdowns in our case. Moreover, in the tracer released method and the combined approach presented in this study, we are interested in the increase of concentrations (the amplitude of the plumes) due to the tracer and targeted point sources above the background signal more than in the absolute value of the measurements. Thus, an offset of the

measured concentrations will not impact our estimates. This will be clarified in the new manuscript.

Line 370-375: The flow meter you refer to is an analog measurement of flow rate, so we don't think you have a time-resolved record of the flow rates. While you have a final check of the release by mass difference, do you have any estimate of the variability of the flow rate over the course of the experimental measurements? This could be substantial, particularly with acetylene releases, which can vary as the cylinder cools. In such a case, the flow rate may appear the same on the flow meter, but the actual mass flow may be different.

We do not have a time-resolved record of the flow rates during the experiments. However each cylinder has been weighted before and after each series of measures and the flow rate calculated with the mass difference was systematically in good agreement with the flow rate read on the flow meter. Therefore we have no reason to believe that there was an important variability of the acetylene and methane release during our experiments. This will be discussed in the new manuscript.

Paragraph starting at 441: When multiple instruments are sampling in a mobile vehicle, they will often have different inlet times (lag due to air intake).

In our case, we use one instrument only and we assume that the inlet time for acetylene is the same as for methane. However, we will now better discuss the fact that, since our instrument measures alternatively methane and acetylene, the measurement times for the two species are slightly different which is a source of uncertainty for both the tracer method and statistical inversions.

This can be corrected based on experimental measurements, for example by delivering an excess of nitrogen or zero-air to the inlet tip, and monitoring the instrument responses. Has this been done experimentally? How does the time-shifting of data based on simulated plumes compare to the experimental lags?

The inlet time of our system has been tested during previous experiments by delivering an excess of carbon dioxide to the inlet and measuring the response time. The estimated inlet time is about 10 seconds. However, as will be better clarified in the new manuscript, the difference between the exact effective wind corresponding to the actual plumes and the measured wind used to force the Gaussian model, and modeled concentrations with inlet time correction does not always perfectly fit with measured concentrations due to slight changes in the wind conditions. That is the reason why we rather use the experimental lag to correct the model-data comparisons for each transect. This will be better discussed in the new manuscript.

3. Why are ideal tracer ratio experiments not producing acceptable results?

Config. 1 tracer release method estimates before Gaussian correction (e.g. Table 2, middle row) overestimate the true release rate by 19%, when it should produce accurate results without the need for correction (see discussion below for more detail). This discrepancy casts doubt on the biases of the remaining 3 configuration tests. The calibration of mixing ratios and the consideration of errors in the instantaneous flow rates (see above) should be considered.

Our explanations is that it is connected to two sources of errors that likely impacted the computations on the first measurements series for configuration 1 than the following ones, and that we under-estimated in the present version of the manuscript:

- the slight differences in time between the C2H2 and CH4 measurements by the same instruments (see above)

- the variations of the CH4, and to a lesser extent C2H2, background concentrations estimated by calculating the standard deviation of both series without the plume corssings (for example 0.3 and 9 ppb for acetylene and methane respectively for configuration 1).

Both raise uncertainties in the computation of the plumes above the background and in the comparison between C2H2 and CH4 concentrations.

These sources of uncertainties will now be better accounted for and discussed. We will update the results to decrease their impact by interpolating the CH4 measurements to the C2H2 measurement time before comparing the plumes, and by selecting another method for defining the background concentrations (by computing their average in parts likely not impacted by the plumes rather than using the 5th percentile) which, in principle, should slightly decrease the sensitivity to their variations. Here is the new table 2 corresponding to this update of the results:

Table 1 – Methane emission rates of the different controlled release configurations estimated with the different approaches and methane fluxes actually emitted during these tests. The uncertainties given with the tracer release method are detailed as follows: total uncertainty (standard deviation of the random uncertainty derived from the variability of the results from one transect to the other one ; bias due to the mislocation of the tracer).

	Configuration 1	**Configuration 2**	**Configuration 3**	**Configuration 4**
Controlled methane release ($g.h^{-1}$)	382 ± 7	428 ± 7	360 ± 7	482 ± 7
Tracer release method estimates ($g.h^{-1}$)	434 ± 23 (0 ; 23)	564 ± 120 (296 ; 416)	321 ± 51 (131 ; 182)	804 ± 160 (352 ; 512)
Relative difference to the control release (%)	14	32	11	67
Combined approach estimates ($g.h^{-1}$)	441 ± 6	358 ± 2	386 ± 2	462 ± 34
Relative difference to the control release (%)	15	16	7	4

Lines 584 – 590: Something seems wrong with this plume analysis. You are well above the detection limit (at least looking at Figures 4 and 5), so the explanation of why the tracer analysis is not accurate does not make sense.

As mentioned above, we will develop the effect of acetylene and methane background variability and of the slight differences in time between the C2H2 and CH4 measurements by the same instruments to explain the difference between our estimate and the actual emission rate instead of the low emission rates.

Configuration 1 should be the ideal case. Please rule out instrument calibration errors (both in measurement and release equipment), unit errors (temperature and pressure of measurement), and other experimental issues.

We will remove the sentence “this misfits associated....” (line 588-590) and explain that the difference between the difference between our estimate and the actual emission rate comes from measurement errors, the difference in time between acetylene and methane measurements, but also from the variations of the background CH4 and C2H2 signal, which cannot be accounted for correctly based on a single “background” value, as in this study. As mentioned previously, such a background computation and the corresponding results will be updated in the new version of the paper to decrease the impact of these variations, and this topic will be discussed.

The accuracy of the sensor used for temperature and pressure measurements is ±0.3 °C at 20 °C and ±1 hPa.

Dual-tracer release has been used in the past to quantify bias. When they are collocated, the emission rate of one tracer as derived from its downwind ratio to the other is found to agree extremely well with the known mass flow (much better than the 19% error that is observed here). See Figure S4-4 of the supporting information for Allen, et al, Proc. Nat. Acad. Sci. (PNAS), vol 110, page 17768 for an example. It is also shown in the attached figure. There, the downwind ratio of two collocated tracers (C2H2 and N2O) is within 1% of the ratio of their mass flow rates (Figure S4-4c). If N2O were replaced by CH4, the scenario in Figure S4-4 would be experimentally identical to configuration 1 of this manuscript. The tracer release-derived CH4 flow rate would then be within 1% of the known flow rate. This level of agreement for collocated sources is routine in the field. Therefore, the 19% deviation observed here strongly suggests that there is an issue with the measurement method (instrumental), or tracer/methane flow rate.

See our answers above.

If configuration 1 were viewed a "control", then the +19% bias indicates that other aspects of the measurements are +19% biased, not the tracer method inherently. In that case, it should be used as an bias offset – that is, the +29% bias for configuration 2 should actually be 29%-19% = 10%, for configuration 3 should be 17%-19% = -2%, and for configuration 4 should be 58%-19% = 39%.

As indicated by our answers above, we do not believe that the 19% reflect a systematic misfit between our known and targeted rate of the CH4 emission and the perfect estimate that we could expect based on our measurement protocol. We rather see it as a random uncertainty associated with random sources of errors in the background concentrations and in the measurements, which should be associated with our estimates rather than corrected for in such estimates (using our knowledge of the true CH4 emissions), and whose impact will not be the same between the different configurations. So we prefer not following such a suggestion (of note is that the discussion on whether error from configuration 1 should be reported to other configurations in term of relative vs. absolute numbers could be very difficult).

Lines 605 – 613: Please better explain why this tracer ratio method is not working, given that earlier you say that position errors perpendicular to the wind should not have such a large effect. Was the wind varying?

During this series of measurements, the wind came from the north-east (there is a mistake in table 1 that will be corrected) and these conditions the shift between the sources is not only lateral. That is the reason why the error due to the mislocation of the tracer is important. We will now better discuss the fact that lateral shifts of the sources perpendicular to the wind direction, in theory, should not impact the results of the tracer release method, even when having several targeted sources, due to the linearity of the estimation, but that configuration 4 did not correspond to such a situation.

Also, the "configuration 4" panel in Figure 5 shows only one methane plume. Given that there are two CH4 sources, why don't you see two CH4 plumes, or at least a broad CH4 plume? We would expect the two CH4 plumes to look like a composite of the C2H2 and CH4 plumes in configuration 3, which has identical separation.

Actually, in Figure 5, as for configuration 1, we did not select a representative figure for configuration 4. It corresponds to a case for which the 2 methane plumes seem to perfectly overlap despite the shift between the two methane sources (we can still notice that the methane plume is wider than the C2H2 one). We will replace this figure by another one showing a clearer separation of the methane plumes (even though they still partially overlap) and the acetylene plume collocated with one of the methane plumes.

4. Initial guesses for optimization

Paragraph at 276 & Line 474: Please explain how initial guesses completely independent of measurements can be done in practice, particularly for emission sources where the expected emission magnitudes may span many orders of magnitude (e.g. factor of 100, instead of 80%). This would be the case, for example, for certain oil and gas emission experiments. Is it possible to use tracer release result (without any dispersion corrections) as the prior?

Such an initial guess for a given industrial site could be given by the product of typical emission factors times for the sector of activity times the level of activity of the industrial site. But yes, this knowledge could be very poor for the examples provided by the reviewers. This is why we use a very high (80%) prior uncertainty in our inversion system. Such an uncertainty, in practice (in our experiments), gives a small weight to the prior estimate in the inversion process, and the inversion results are actually very weakly sensitive to the choice of this prior estimate. It will be discussed in the new manuscript.

The statistical inversion framework assumes that the uncertainties in the prior estimate, in the model and in the observations are fully independent, which would not be the case if the prior estimate is based on the tracer and methane measurements which respectively feed the model and are used as observations for the statistical inversions. We would definitely account twice for such data if using the results from the tracer method as a prior estimate of the inversion.

Specific Comments:

Lines 191 - 196: Other methods of calculating the ratio are commonly used, notably taking the slope of the plot of the methane vs tracer plume signals. This is generally found to be more precise than measuring the area under each plume, because it does not depend upon choice of background. See Roscioli et al: 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.; et al. Measurements of methane emissions from natural gas gathering facilities and processing plants: measurement methods. Atmos. Meas. Tech. 2015, 8, (5), 2017; DOI: 10.5194/amt-8-2017-2015.

We will mention this option and cite this paper when discussing the options for the observation vector.

Lines 225 : There is considerable debate in the atmospheric modeling community on the timescale appropriate for the canonical A-D stability classes (e.g. Pasquill- Gifford stability classes). The consensus seems to be more on the order of 10-15 minutes than 1-2 minutes. See, for example: Fritz, B. K.; Shaw, B. W.; Parnell, C. B. J. Influence of meteorological time frame and variation on horizontal dispersion coefficients in Gaussian dispersion modeling. Trans. ASABE. 2005, 48, (3), 1185; DOI: 10.13031/2013.18501

We agree that the selection of the stability class per transect as a function of the continuous wind measurements based on such a table is not perfect, which is why we would have been ready to use a class of stability that does not fit with the wind speed (according to the Pasquill stability classification) in order to get the best fit with C2H2 data as possible. However, it appeared that the stability classes that provided the best fit with C2H2 data were systematically part of those corresponding to the measured wind speed. This will be better discussed.

Line 302: ". . . spatial offset between the measured plume and the actual plumes due to the lag between the air intake and the concentration measurement." I recommend adding the parenthetical statement: (also known as inlet lag or inlet time)

We will use this term in the paper every time the topic will be discussed.

Line 340: correct "serie" to "series"

This will be corrected.

Line 351: While the instrument reporting time is noted (2 seconds), the instrument response time is not. Furthermore, the data depicted for configuration 1 in Figure 5 suggests the time response for the two channels is not fully matched. Is the instrument reporting all mixing ratios simultaneously, or does the instrument sub-divide the 2 second interval to quantify each of the noted species.

As mentioned before, we will be better discussed the inlet time of our system in the new manuscript. The instrument does not measure both species at the exact same time and there is actually more measurements of acetylene than methane (methane is measured approximately every 2 seconds and acetylene every seconds). This will be better described in the analytical equipment part. In particular, as discussed above, we think that this interval between the acetylene and methane measurements is a significant source of uncertainty in our calculations.

Line 429: Is the 5th percentile of transect concentrations sufficient to determine baseline? Do the results change if the 2nd percentile, or 10th percentile is used?

The definition of the background (or baseline) will indeed impact our estimates and we assume that it is one significant source of uncertainty in our experiments (see our answer to comments on configuration 1 above). We now think that taking a specific percentile (i.e. a given data) was not the best option for this. This is why we have updated our results with a new background estimated as the average of the concentrations measured before and after the plumes and the new manuscript will discuss this topic of the sensitivity to the definition of the background.

Lines 561 to 567 and Figure 3: Why not show these results as a function of relative distance downwind, i.e. (distance between source and tracer)/(distance between site and measurement)? The distances in meters shown cannot be easily generalized. I further suggest putting panels a) and b) on the same vertical scale, and adding gridlines at the same intervals for both.

We agree with this suggestion but we will still provide both relative and absolute distances since the actual atmospheric transport processes will change depending on the absolute distance so that we should be very cautious if generalizing such results. We rather take it as an illustration of the amplitude that the bias can reach in a rather simple experimental case.

All Figures: Please increase font size of all labels and numbers so that they are readable at the width of a normal sheet of paper.

OK

Figure 4: Consider showing only Figure 5 (representative plume transects) instead of the whole data set. These full data results, depicted in Figure 4, might be better left to the supplementary info. Even better, consider publishing these results as a test dataset. If Figure 4 is to remain, then the vertical axes should be rescaled to see all of the plume intensities.

We would like to keep Figure 4 to illustrate the variability of the background, especially for the methane. We agree about removing the highest measured concentrations that do not correspond to actual plume crossings, but to measurements close to the cylinders when we checked the stability of our emissions. These concentrations were so high that the actual plume crossings could not be seen properly.

Figure 6: Figure 6 is the most important figure of the manuscript and needs to be reformatted for

legibility. It is currently much too small. Consider perhaps a small cartoon drawing of "config. 1" and "config 4". To clarify the difference between the 2 sets of results shown. Please also label the meaning of the shaded area ("concentrations" plots) in the legend or figure title.

We will improve this important graph by increasing labels size, better indicate the difference between the top and the bottom graphs and explain that the shaded areas correspond to the uncertainties of the modeled concentrations.

Line 565: Instruments with lower levels of detection than the one used here are available. Please alter this statement to reflect this fact: e.g.: ". . .signal to measurement noise ratio would likely be too small, using these instruments, to derive. . ."

We will follow this suggestion.

Section 4.3, Lines 614: Please reference Goetz et al, who in 2015 used a similar approach to this.

As mentioned previously, we will cite Goetz et al. in the section 2.2 and 2.5.

Line 651: It is possible, (see Roscioli et al.) in some cases, to co-locate tracers and emission vectors for real experiments. Please rephrase "which can hardly occur"

Fugitive emissions (like leakages) that are transitory, or affecting poorly reachable areas or complex buildings, unexpected sources, but also widespread and heterogeneous sources (e.g. livestock in the building of a farm, basins in waste water treatment plants or cells in landfills for which emissions are not homogeneously distributed) makes it difficult to know perfectly the distribution of the emissions within a site. This sentence will be modified and extended to clarify it. In particular we will replace "which can hardly occur" by "which is not always easy in real cases".