Response letter to Review #1

Responses from the Authors are given in blue italicized text throughout: Thank you for your helpful comments. We addressed them to the best of our understanding and appreciate your guidance on areas where you found our discussion or language unclear. Items such as font size, word choice, and grammar which will be fixed in the revised manuscript are noted with gray highlighting.

Summary

This paper presents a methodology that estimates CO_2 and CH_4 fluxes using cylindrical flight patterns combined with kriging and Gauss divergence theorem over Sacramento, California and quantifies the corresponding uncertainty. The study finds that fluxes vary as a function of wind pattern, seasonality, background assumption, flight path, and flux estimation approach by a factor of 1.5 to 8. Total flux estimations using the entire circumference are larger than if just downwind region is used. It is stated in the article that using entire circumference to estimate GHGs fluxes allows for accounting of unknown sources that otherwise could be missed.

General Comments (Major Revisions)

Although the paper does make a lot of important and useful points regarding estimation of GHGs emissions with an aircraft, there are many places that are unclear and need to be elucidated before I can accept this article for publication.

First, I am not exactly clear on how the used methodology is different from a traditional mass balance. I see the explanation, but I am not convinced that it provides any information that is not obtained from the standard method. I would like to see the comparison. Please perform standard mass balance and compare it to your method.

From the Authors: The basic approach we used for estimating CO_2 and CH_4 fluxes is not different from a traditional mass balance. In the traditional mass balance analysis, the incoming mass should be the same as the outgoing mass passing the X-Y plane of the measurements. So, the air pollutants from the city will be observed on the downwind side, along with the wind passing through the region. Similarly with a cylinder of measurements, the mass coming into the cylinder should be equal to the mass leaving the cylinder to satisfy the mass balance.

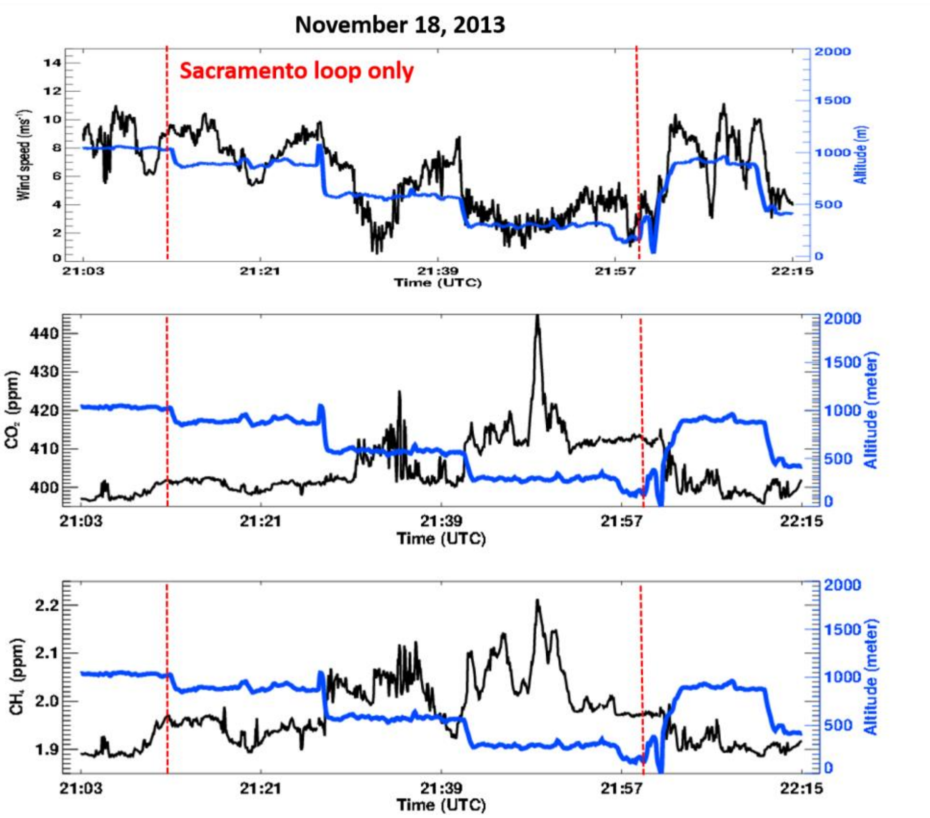
When we used the raw wind measurement, however, we found that the mass of air mass was not conserved, which means we may not fully apply the mass balance approach. That is why we used the mass-balanced mean wind at each grid level, and the mean wind will distribute the greenhouse gas inside the cylinder. We assume that the well-mixed condition applies.

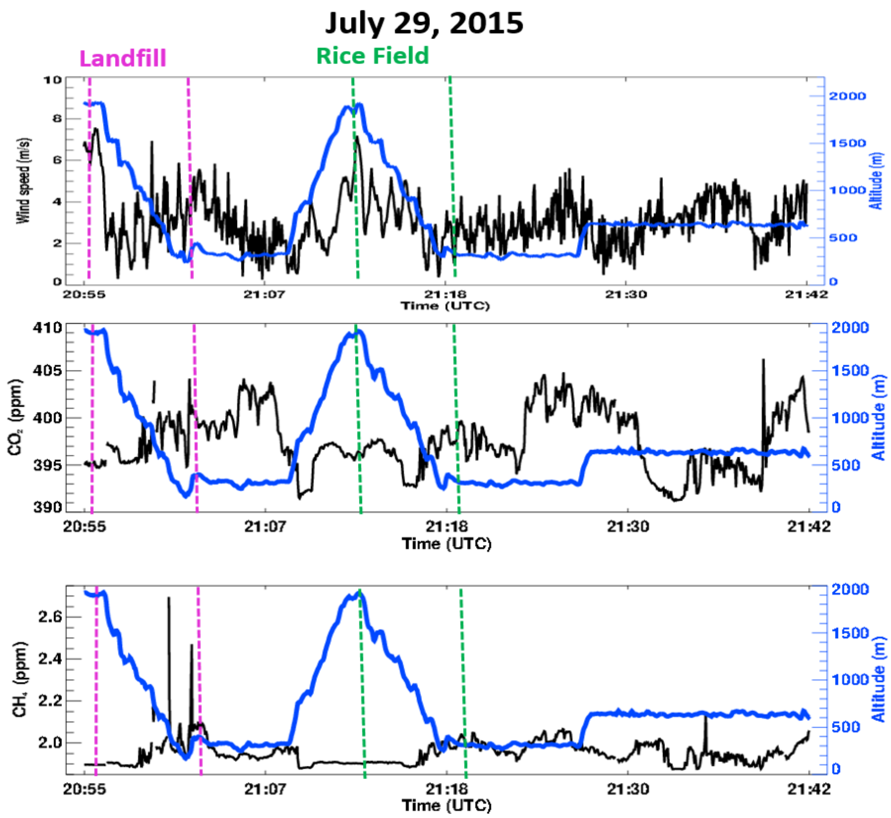
Furthermore, one of the advantages of applying the mass balance approach with an oval (enclosed shape) flight path is that an assignment of the background concentration is not required. Because we calculated the mass of the GHG plume that entered and exited the sample volume, the background effectively cancels out. Sometimes (actually almost always) background is not homogeneous, so there can be large uncertainty associated with calculating the background. There are also challenges in isolating the plume. In our approach, such problems did not exist.

Our study is also different from the ususal implementation because we used in situ measured winds, not model data. Furthermore, we used the mean calculated from measured wind at each level, not just one *level, as has been the case in many previous studies (Turnbull et al., 2011; Karion et al. 2013, 2015). The difference we adopted here is that 1) we tested the flux estimation using observed airborne high-resolution raw wind at each measurement point and kriged grid and 2) we tested the flux estimation using mean wind (averaged observed raw wind at each layer). The previous "standard mass balance approach" often adopts the wind data from coarse-resolution model output. The uniqueness that our study provides is that we showed how the final flux could be different depending on the wind treatment.*

Another issue that I find in this article is that the actual plane data is not carefully presented. First, it is important to present data for all of the 3 cases in an equal manner. There are different plots for different days, and it becomes confusing. For example, finding the exact local time of all the flights is difficult (including just take off time is not enough in this case). This information needs to be easily accessible. I could not locate wind measurements for all of the days. Figure S2 (d) is misleading and leads to a flawed assumption regarding steady state for November 17, 2015 (more on that later). The paper needs to be reorganized and improve its clarity of presentation.

A time series plot of CO_2, CH_4, wind, and altitude for each analysis day is shown below and will be included in the revised Supplemental Information.
In addition, we will amend the text in Section 2.1 to focus on the time during which sampling occurred and simplify the way we described the implication of daylight savings time in California (PDT vs. PST). Revised text will read: Sampling occurred 21:10 – 22:00 UTC for the November flight (local standard time is UTC minus 8 h, 13:10 – 14:00 PST) and 20:55 – 21:45 UTC for the June flight. Based on comments from both reviewers, we have removed the November 17, 2015 case. The revised manuscript has changes to figures and the table to reflect this removal.





The explanation of background is very confusing. Given the method presented, background should be everything that flows into the cylinder. I am not following the justification for different background assumptions (you would not want to pick a minimum value in this case).

Yes, ideally, the cylindrical flight design removes the need for assigning a background value, as you state. Because we calculated the mass of the GHG that entered and exited the cylinder, we are subtracting all GHG upwind from all GHG downwind. So, the background should cancel out.

We tested this expectation by calculating the flux using various wind and background treatments. In the linear curtain approach, the edges of the flight transect outside the plume or measurements made upwind usually provide the background values. In these cases, the background concentration can be non-homogeneous and difficult to specify, so there is large uncertainty associated with setting the background. Because we want to compare as directly as possible our method to the studies already in the literature, we started by assigning the background as the minimum value in the layer (as a parallel case to the "edges" method). Then as a sensitivity test, we explored how the results would or would not change with a different choice of background (average of the layer values). When the winds were used in a way which required mass balance ("mean wind" in Table 1), the choice of background value was shown not to matter (first and second lines of Table 1 and Table 2). The "raw wind" case is discussed in detail in the next item.

We hope the new structure of the manuscript will better communicate this approach.

Also, the concept of raw and mass-balanced mean wind needs to be better explained. Why averaging winds horizontally achieves mass balance? And if plume is not well mixed, how can you do that? Plume is transported differently at each level. You cannot just assume that all of the levels move at the same rate.

The mass-balanced mean wind is the arithmetic mean of the inflow and outflow raw (measured) wind at each vertical level. Separate calculations were performed to test the sensitivity of the calculated flux to the thickness of the vertical levels. These results will be incorporated into an expanded version of Table 1 in the revised manuscript. This table can be seen below. We will expand it in the revised manuscript, separate it into two tables [Table 1 (urban scale), Table 2 (local scale)], and also include a new Table 3 showing comparison with the Turnbull et al. study and bottom-up inventories.

Table 1. Urban scale fluxes over Sacramento on November 18, 2013.

Background	Wind		Urban Scale (large loop) November 18, 2013 CO_2 (Mt yr^{-1})	CH_4 (Gg yr^{-1})
min	Mass-balance	100 m layer avg	25.6±2.6	87.1±8.7
		500 m layer avg	26.6±2.7	88.7±8.9
		Whole column avg	26.6±2.7	88.7±8.9
avg	Mass-balance	100m layer avg	25.6±2.6	87.4±8.7
		500m layer avg	26.6±2.7	89.0±8.9
		Whole column avg	26..6±2.7	89.0±8.9
min	Raw		3.7±0.4	13.0±1.3
avg	Raw		25.5±2.6	91.1±9.1

Table 2. Local scale fluxes over landfill and rice field over Sacramento on July 29, 2015.

Background	Wind		Local Scale (small loop): July 29, 2015			
			Landfill		Rice Field	
			CO_2 (Mt yr^{-1})	CH_4 (Gg yr^{-1})	CO_2 (Mt yr^{-1})	CH_4 (Gg yr^{-1})
min	Mass - balance	100 m layer avg	0.2±0.1	7.1±2.5	0.3±0.04	2.6±0.4
		500 m layer avg	0.2±0.1	7.1±2.5	0.3±0.04	2.6±0.4
		Whole column avg	0.2±0.1	6.9±2.4	0.3±0.04	2.6±0.5
avg	Mass-balance	100 m layer avg	0.2±0.1	7.1±2.5	0.2±0.04	2.6±0.4
		500 m layer avg	0.2±0.1	7.1±2.5	0.2±0.04	2.6±0.4
		Whole column avg	0.2±0.1	6.9±2.4	0.2±0.04	2.6±0.4
min	Raw		0.4±0.1	9.0±3.1	0.2±0.03	1.7±0.3
avg	Raw		0.2±0.1	7.1±2.5	0.3±0.04	2.6±0.5

Table 3. Flux estimates for the Sacramento urban area from measurements made on November 18, 2013.

		CO2 (Mt yr^{-1})	CH4 (Gg yr^{-1})
Whole cylinder–AJAX	(bg = min, 100m layer avg)	25.6 ± 2.6	87.1 ± 8.7
	(bg = avg, 100m layer avg)	25.6 ± 2.6	87.4 ± 8.7
Curtain –AJAX	(bg = min)	17.3 ± 1.7	64.4 ± 6.4
	(bg = avg)	8.9 ± 0.9	24.1 ± 2.4
Turnbull et al. (2011)		13.6 (with uncertainty of ~ 100%)	
Vulcan estimates for Sacramento		11.7	
CEPAM estimate for Sacramento		10.3	

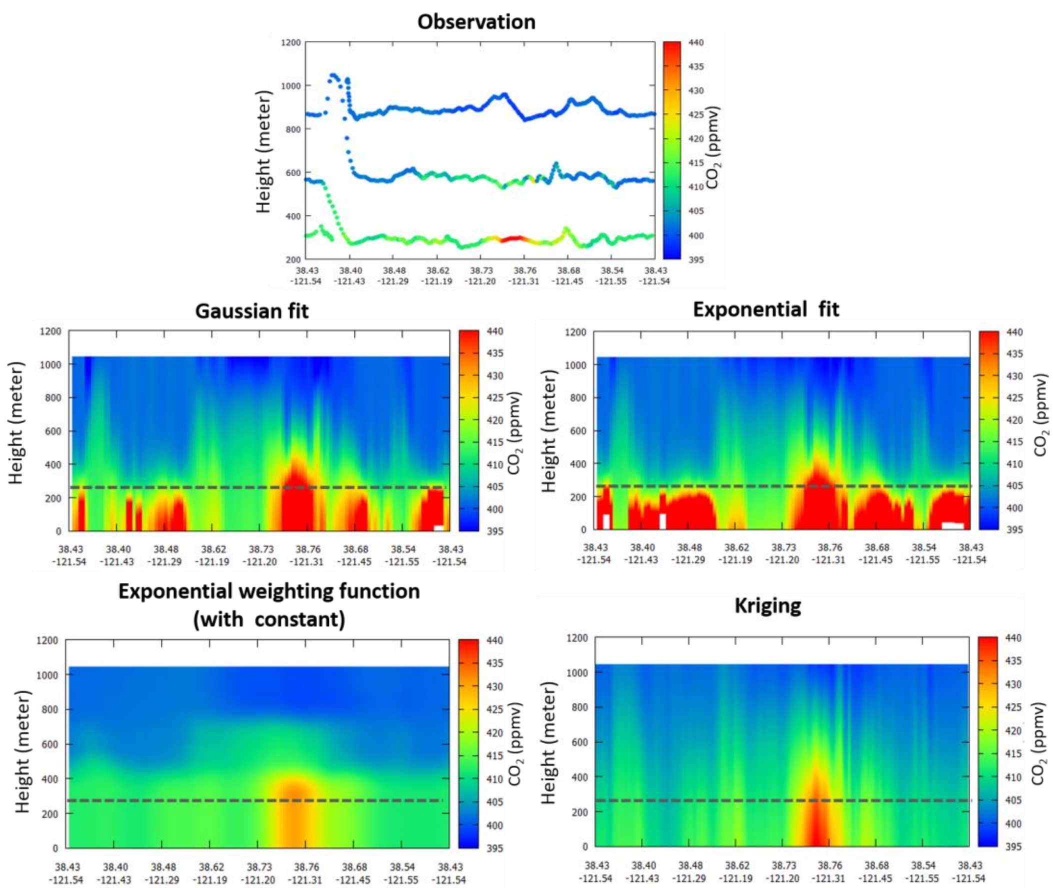
[a] Turnbull et al. (2011) data was collected in 2009; the value given here was converted from the mean reported value of 3.5 Mt C yr^{-1} with a 1.1% yr^{-1} increase in CO_2 flux to adjust to 2015.
[b] Bottom-up inventory estimates of the annual total emissions from Sacramento County from Vulcan (Gurney et al., 2009) and the California Air Resources Board CEPAM database (Turnbull et al, 2011) are included for comparison. The Vulcan inventory is available only for 2002, and the CEPAM database is available for 2004. We applied a 1.1% yr^{-1} increase in CO_2 flux to adjust to 2015.

Previous studies used a single coarse resolution model 'mean' wind throughout all altitudes below the PBLH (Turnbull et al., 2011; Karion et al. 2015), and we agree with you that assuming the winds at all levels move at the same rate is not ideal. Therefore, we used our in situ wind measurements to test the impact of this assumption. We calculated the fluxes using wind averaged on each vertical layer (three separate tests with layer thicknesses of 100, 200, and 500 m) and found the results were actually not very much different from the flux estimate found when we use a single, whole column average.

Specific Comments

Line 170: It says, “The background level is derived from the lowest flight measurement.” When using the kriging method, do you apply kriging to all of the data including the background?

We apologize for the wording in the original manuscript at line 170. Our description of the "constant" method was unnecessarily confusing. The blue curves in Figure 2 show what we were trying to describe: at all altitudes below the yellow diamond (lowest flight level), the mixing ratio was presumed to be exactly equal to the value measured at the lowest flight level. Please also see the new figure below. The bottom left panel shows more clearly how the values measured along the lowest flight level are assigned to all grid cells below the measurement altitude. The dashed lines represent an approximate lowest flight level.



Line 191: How do you know that kriging approach captures better plume features? Kriging method interpolates data, meaning that it basically guesses it. It has no knowledge of the actual plume dispersion mechanism. The Figure S4 is misleading as it has different color bar scales for different plots. Please make sure that all of the color bars are the same. In actuality, you don’t know how the plume is changing below your lowest measurement point. Anything that you assume below that point is highly uncertain. It could be almost constant for all we know. It really depends on the location of the source.

That’s why ideally you want to sample the well-mixed layer and not partially mixed layer with an aircraft when estimating flux.

That is exactly why we evaluated several common methods of extrapolation, as shown in Figure 2 and the new figure above. As you point out, we do not know which method will better capture the plume features, especially below the lowest flight level. Because kriging uses characteristics of the measured data to make the interpolation, we expect it to be less arbitrary than some of the other methods, and this is what is seen in figure S4. The measured plume shape (panel b) is captured more faithfully by the kriging method (panel c) than by the exponential method (panel d), as is the subtle variation of mixing ratio around the oval at a given altitude.

We agree with you that plume behavior below our lowest measurement point is highly uncertain. We don't think one pattern is correct while others are not. We just suggest one estimate might be better based on the other characteristics that we've examined. Your comments on why a well-mixed layer condition is important for estimating flux are also very true.

More attention will be paid to the color bar scales in the revised manuscript, and where appropriate they will have the same scale, such as panels b, c, and d of Figure S4. However, if forcing different flight days to use the same scale makes it more difficult to discern features of interest, then we will not force one day's maximum value to set the scale bar for another day's flight. If different scales are necessary for clarity, we will note the change in the relevant figure caption(s).

Section 2.4: See the comment about the raw wind vs. mass-balanced mean wind in the general comments section.

Please see reply above.

Figure 1c: I am confused about the following sentence in the caption, "The shading represents the pressure . . . normal to the cylinder." What shading? I am not sure I see any shading. Please explain what do you mean here. Also, here you say that blue is inflow and red is an outflow. It seems that everything that is in blue should be a background for everything that is in red assuming steady state. Please comment.

We will change the word choice in the revised manuscript. "Shading" was intended to mean "colors" of the cylinder. This is just the sign of the air mass flux [kg m^{-2} s^{-1}], which is obtained from density multiplied by the wind vector (or pressure divided by the wind vector). Yes, everything in blue represents the inflow air mass, which has negative wind direction, while everything in red represents the outflow air mass, which has positive wind direction.

Lines 265-270: I do not understand your choice of background. Given your set up you should be using inflow as background. The definitions you describe here are used in regular mass balance because sometimes there is just not enough sampling, but generally, they are flawed. Please explain why you are not using inflow. You need to justify your choices with relevant physical processes.

We agree with you that we do not need to know the background value for estimating flux when adopting the circular pattern of flight. This is clearly demonstrated in our experiments, showing that the flux estimates were not sensitive to the choice of background (minimum or averaged value). We will work harder on the language in the revised manuscript.

Another important point that you do not mention is an uptake of CO_2 by vegetation. That also can affect background and your fluxes quite a bit. I know it is November in two of your cases, but you need to comment on your assumptions. Your case in July could be more problematic with respect to CO_2, although there you concentrate on CH_4 so it may not matter as much.

We agree that considering an uptake of CO_2 by vegetation could be an important factor on total flux estimate. When we took a look at CarbonTracker data, we confirmed that there is some contribution of the vegetation of CO_2 to the total fluxes in November (for example, in places like Salt Lake City). However, it is hard to consider the biological impact on CO_2 flux unless we downscale the model data to the small scale we are interested in. This would be better considered in a further study to completely characterize each sector (biological (vegetation or dairy farm) or anthropogenic (industry)). However, we agree with the reviewer that we should mention the potential problem in the interpretation of the flux estimate we obtained in this study and will include a comment in the revised manuscript.

Line 306: How come highways and airports are indicative of CH_4 emissions? It is not common for these sources to emit any significant CH_4. Please explain.

Dairy farms and landfills are well known sources of CH_4. However, one of biggest concerns regarding CH4 emission is the contribution from unknown sources. We see the slight increase of CH_4 emission over those sites. Broken pipe lines or other facilities at the airport could be possible sources of CH_4, so we called it to attention. But, as you pointed out, this cannot be a deterministic source of CH_4. We will modify/rephrase the sentences in the revised manuscript.

I think using kriging when you do not understand your sources is a risky endeavor. It is better to solve for everything without kriging first and then see how kriging may affect your results. But in your situation, you definitely do not want to trust kriging. Using kriging in regular mass balance is also dangerous if you do not have a good understanding of what you are measuring. Unfortunately it is often used without much thought. For example, see Figure 6 in Conley et al. (2017), the paper also uses the divergence methodology that you apply here, but they are careful to note that you want an optimal number of loops around your source before you can get a stabilized estimate of emissions. They estimated an optimal number of loops to be about 15 to 25. That is the case because turbulent conditions tend to increase the magnitude of the random error. I am afraid your sampling here is just too small for a good application of divergence theorem. It is important to acknowledge it. Solve without kriging and see what you can get.

We understand your concerns, and we appreciate the "riskiness". When we calculate fluxes based only on the measured data, without filling in the gaps between flight levels, the total flux estimate will obviously be much smaller than when we account for the entire surface of the cylinder using interpolated data. With an urban-scale cylinder (with a circumference on the order of 100 km), it is impossible to map out the entire surface (~100 km^2) with dense measurements. Although kriging cannot be better than actual observations, it can be a good alternative to “mimic” actual data. We disagree with the reviewer's opinion that we solely rely on the kriging without an understanding of the data. We carefully performed the variogram analysis, and carefully chose the kriging parameters (sill, range, and nugget) based on the experimental and theoretical variogram obtained from the actual data we measured.

Figure 2S (b and d): You will have to eliminate November 17, 2015 case from your article. You cannot assume steady state conditions on a day with calm to variable winds near the surface. The wind rose is misleading as you mainly show free tropospheric winds, which should not be used for boundary layer flux calculation. Your boundary layer winds have no consistent direction. The data from a local weather station in Sacramento, CA supports that (and actually if you look carefully at your wind data you will see it too in your Figure). This comes back to the point I made earlier, where you need to show your actual wind data from every case. You cannot just pick and choose what you show. It is no surprise that your flux estimations did not work well on that day. None of the aircraft methods would work on that day. It is very important to have a good forecast before you go and fly a mission of this type. I am not sure who designed this flight and for what, but it does not work here for your purpose. Perhaps you can find another flight that works better.

Done. Please see discussion above.

Reference

Conley, S., I. Faloona, S. Mehrotra, M. Suard, D. H. Lenschow, C. Sweeney, S. Herndon, S. Schwietzke, G Petron, J. Pifer, E. A. Kort, and R. Schnell: Application of Gauss's theorem to quantify localized surface emissions from airborne measurements of wind and trace gases, Atmos. Meas. Tech., 10, 3345-3358, https://doi.org/10.5194/amt-10-3345-2017, 2017.

Karion, A. et al.: Methane emissions estimate from airborne measurements over a western United States natural gas field, Geophys. Res. Lett., Vol. 40, 1-5, doi:10.1002/grl.50811, 2013.

Karion, A. et al.: Aircraft-Based Estimate of Total Methane Emissions from the Barnett Shale Region, Environ. Sci. Technol. 2015, 49, 8124-8131, DOI: 10.1021/acs.est.5b00217, 2015.

Turnbull, J. C., Karion, A., Fischer, M. L., Faloona, I., Guilderson, T., Lehman, S. J., Miller, B.R., Miller, J. B., Montzka, S., Sherwood, T., Saripalli, S., Sweeney, C., and Tan, P.P.:: Assessment of fossil fuel carbon dioxide and other anthropogenic trace gas emissions from airborne measurements over Sacramento, California in spring 2009. Atmos. Chem. Phys., 11, 705–721, 2011, doi:10.5194/acp-11-705-2011, 2011.