Response to Referee # 2

We would like to thank the reviewer for their helpful remarks. Below we address all of the comments presented to us by the reviewer.

There are a few sections which may require some clarification, and it may be helpful to the reader if the authors reordered some of the content.

The order of the subsections in section 2 (Methods) does not follow the order that the four components of the inverse modelling framework are listed in the abstract or discussion section. It may make it easier for the reader to follow from the methods through to the results if there is an explicit methods section for each of the four sensitivity analyses.

The sections are now slightly reordered to better reflect the whole structure of the paper.

Section 2.4 lines 172 to 185: Could the authors explain what the percentages are referring to and provide the overall quantity? Does this come from an inventory analysis?

These are percentages of the total Indianapolis methane emissions depending on a given estimation. For instance, some estimates (Cambaliza et al., 2015) found landfill to contribute only to 22% of the total Indianapolis methane emissions, while other estimates (GHG reporting program) found landfill to be responsible for 63% of total city methane emissions. The total emission value could be the same or different. Some estimates come from inventories; others come from top-down studies such as aircraft mass balance and inversion. Figure 1 in the article summarizes different estimates of methane at Indianapolis. It just does not have the breakdown of the total emissions by sources. We know the breakdown only for some studies. There is now an attempt to summarize this breakdown in section 2.7.

The section following, starting at line 187, is also labelled as Section 2.4.

This is an error. It is now fixed.

Section 2.5 line 218: Should this be a subsection of the previous background section (2.4)? Or should this subsection be called Variability in Background Concentration?

Section 2.5 describes bivariate plots, so it is titled correctly. Due to the reordering of the methodology part of the paper some sections now changed their number.

Equation 3 line 250: This equation and notation are taken from Hanna et al. 1982. It may be useful to the reader to explain the terms more thoroughly. Most of the explanation for terms is taken from Chapter 9 of this text, but some of the terms are explained more fully in earlier chapters. I was particularly distracted by the zi until I realised that it was conventional for the zi to denote the boundary layer height of the box, rather than i as an indicator for height at time step i (or distance i). Do you take zi to be the average boundary layer height during the hour?

Yes, zi is confusing variable name, so we have changed it to H. Yes, we assume H represents boundary layer height for a given hour. We are thinking to leave the rest of the description as is (some minor edits are incorporated), but if there is anything that you would like us to explain specifically we could do that.

In the text the units of Qa are described to be in mass per unit time per unit area, and the units of C are described to be in mole fraction in the first instance. Should the concentration not be converted to moles per volume? Later on from line 263 this does appear to be the case.

You are absolutely correct. It was an error; it should say concentration in the first instance. On line 263 it states that because our data is given in mole fractions we must first convert it to concentration (mass per volume) before we can apply the budget equation. Correct, we are not converting CH4 to moles per volume, but to complete abovementioned conversion from CH4 mole fractions to CH4 concentration we do need to use average molar dry density of air, which has units of moles per volume.

Section 3.1 lines 292 to 307: This information seems to be related to methods, and should perhaps be in Section 2, probably before the methods section on Background concentrations to be consistent with the abstract and discussion sections.

Lines 299 to 307 have been moved to methods as requested. The other lines seem to fit appropriately to the results section as they give the domain issue some context. It would be to jarring to jump into line 308 right away.

Line 350: I think it would be useful to justify why you needed to use the bootstrapping approach. Lines 350 to 363 and Figure 6: I would propose that the analysis for assessing the number of measurements required to obtain a reliable background concentration estimate should focus on the standard deviation rather than on the standard error.

Bootstrapping approach allows us to vary sample size of a theoretical experiment and to see how the background uncertainty responds to it. Below I explain why this may not work for the standard deviation.

It is unsurprising that the standard error (standard deviation of the mean) reduces as the sample size is increased, as you state in line 359. This is a property of the standard error. If you're interested in the stability of the background concentration difference estimate, you should rather be looking at the standard deviation of the concentration differences (which you can do in the same way, except instead of looking at the standard deviation of the mean values for the 5000 samples, you look at the mean of the variances for the 5000 samples – which can then be converted to a standard deviation). The standard deviation will provide information about the uncertainty in the background concentration. The plots in Figure 6 are slightly different for each wind direction because the standard deviation of the background concentration from these directions differs and because the bias for each direction differs. At some optimum sample size the standard deviation of the concentration differences will stabilize for each direction. If you assume that under normal circumstances (where you don't have two competing domains or multiple background sites) you would be able to obtain the uncertainty in the background concentration from historical data and this uncertainty remains stable over time, you would want to ensure that the background concentration estimate is obtained from a sufficiently large sample size so that you know what the uncertainty in this estimate is. Therefore, if you have multiple background sites where you can assess at which sample size the standard deviation of differences between background sites stabilizes, you may want to determine for each wind direction what this sample size is and what the stabilized standard deviation is. Basing the background concentration on a sample of this size or larger should provide an estimate with a predictable uncertainty, which is now independent of the sample size.

If you know what the standard deviation is, then it follows what sample size is required to obtain a background concentration with the required standard error (precision) (if you can use SE = SD/sqrt(n)).

If you decide to stick with the standard error plots, I think you should show on each plot at which sample size each direction reaches the required precision, say 3 ppb, as this differs for each wind direction.
Thank you for such a thoughtful comment. This really gets me thinking deeper about this background variability concept.

I agree with you that it is unsurprising that standard error is decreasing with increasing sample size. However, I do find Figure 6 useful in that standard errors vary depending on wind direction. So given a sample size of around 20 we can get a general idea which wind direction will provide us with the lowest random error of the difference between backgrounds. It is important to note that this Figure 6 does not say anything at all about bias. This is just random error. The bias is a difference between backgrounds and is shown in Figure 5. In ideal world we would want this difference to be zero and have very little random error on it. Unfortunately that is not the case here. The background is quite complex. Now, you do make a valid point regarding standard deviation and that it also gives us information regarding background variability. Because of this I added to Figure 5 another plot that shows standard deviation times 2 for each of the wind directions. That shows us potential background discrepancy that can occur on a given a single day. This is also useful. Here is the updated Figure 5. On the other hand, standard error shows us that as sample size increases our average difference of backgrounds would approach a known bias. But yes, on any given day things could be really variable or not so variable. Additionally, standard deviation plot indicates that W is the best direction regarding the background. It has the lowest variability of background differences. It does have a bias, but overall error is the smallest. This also is evident in Figure 6, where W standard error is the smallest.

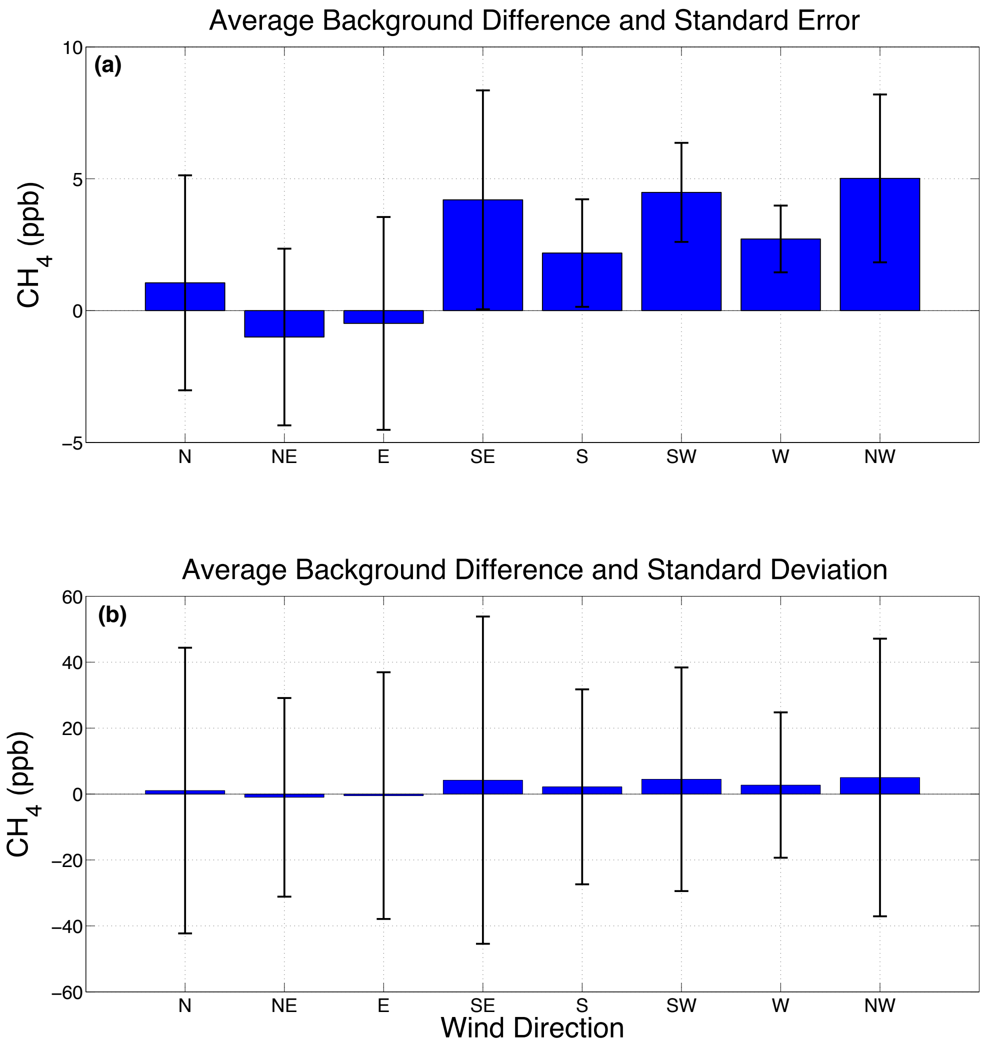

Figure 5. Average of the differences between criteria 2 and 1 CH_4 backgrounds at Indianapolis as a function of wind direction. These averages are generated from the same data as used in Fig. 4 and reflect results shown in Fig. 4g. Error bars indicate in (a) standard error × 2 and in (b) standard deviation × 2.

To respond to your other request, I tried to perform the bootstrap experiment with standard deviations, but it does not seem to work. The variance stabilizes very fast and does not seem to be a function of sample size (maybe only initially). So I think that it would suffice to add a standard deviation plot to Figure 5 because these standard deviations are basically the same ones you would get with the bootstrap experiment. I think this happens because we are sampling from the pool of the same differences and ultimately there is no way for variances to change much after 5000 iterations. In other words, low sample size with 5000 iterations will be similar to a large sample size with 5000 iterations because both of these cases sample from the same PDF of background differences. Please see Figure 6b.

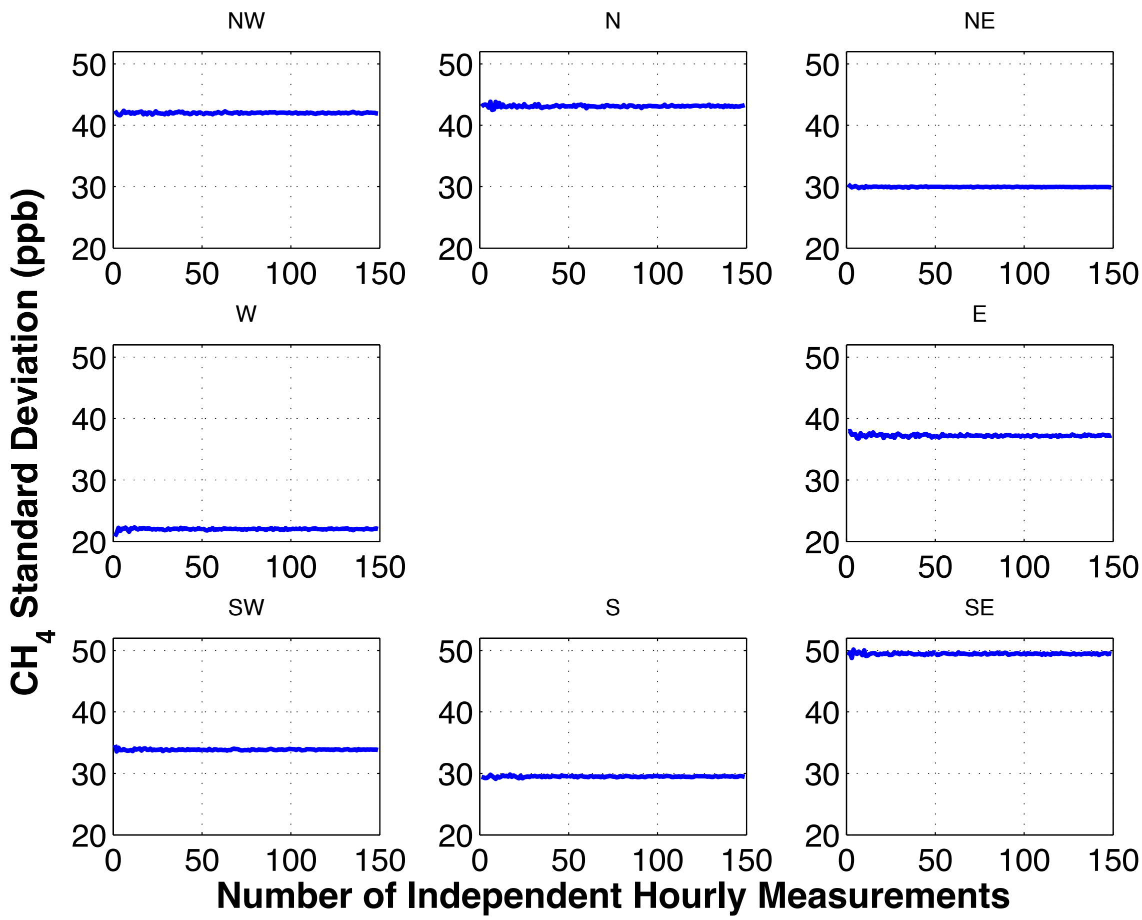

Figure 6b. Bootstrap simulation of the standard deviations multiplied by 2 in Indianapolis CH_4 background mole fraction differences (between criteria 2 and 1) as a function of sample size and wind direction (see text for details).

To answer your last part of this question, I do not think that we have a required precision at this point for background differences. The plot illustrates at approximate what sample size we would approach a reasonable standard error. Perhaps one way to classify a required precision would be to make sure that the standard error (random error) reaches a point where it is less than Indianapolis enhancement of about 12 ppb (a higher estimate of the Indianapolis enhancement from section 3.2) by a factor of 2 when combined with a bias (in this case it is helpful to think in terms of absolute magnitudes, so let say the requirement is 6 ppb). In this case each wind direction would have a different threshold. For instance, given this requirement NW direction would need a random error of 1 since its bias is 5. For NW direction, this threshold would require more than 150 samples. For N direction on the other hand, where the bias is 1, the requirement is fulfilled when random error crosses 5 ppb at 74 samples. This is going to be added to this to the paper now.

Here are the absolute values of the biases for each wind direction, their respective required thresholds, and needed sample size:

N: 1 ppb	T: 5 ppb	N: 74
NE: 1 ppb	T: 5 ppb	N: 36
E: 0.5 ppb	T: 5.5 ppb	N: 46
SE: 4 ppb	T: 2 ppb	N: >150
S: 2 ppb	T: 4 ppb	N: 53
SW: 4.5 ppb	T: 1.5 ppb	N: >150
W: 3 ppb	T: 3 ppb	N: 52
NW: 5 ppb	T: 1 ppb	N: >150

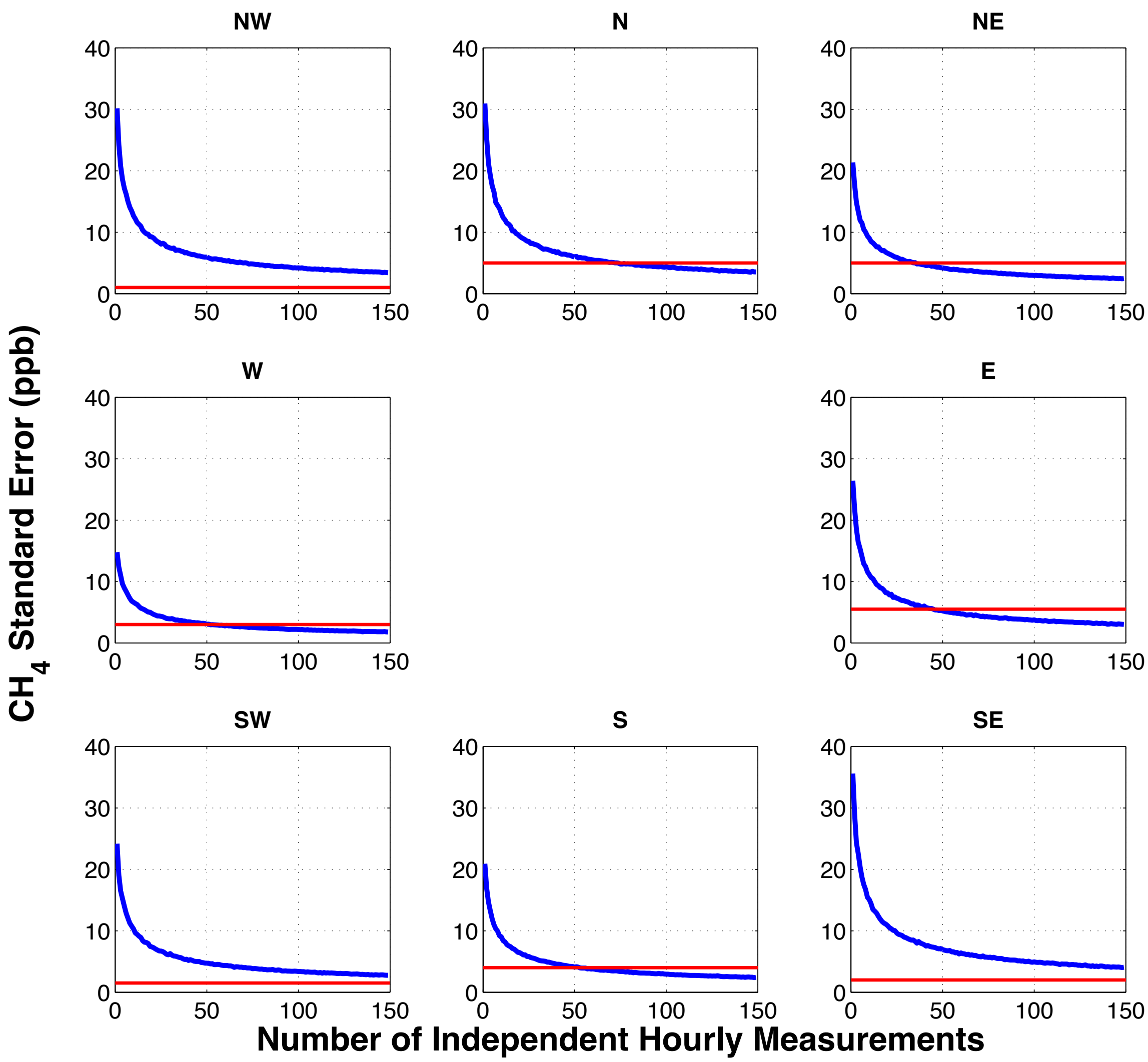

Figure 6a. Bootstrap simulation of the standard errors multiplied by 2 in Indianapolis CH_4 background mole fraction differences (between criteria 2 and 1) as a function of sample size and wind direction (see text for details). Thresholds for each of the wind directions indicate a random error threshold needed for the background uncertainty to be within 50% assuming average CH4 enhancement from Indianapolis is 12 ppb.

A point that should be discussed is that the measurements you obtain for the background site are taken at different times, and as the number of measurements increases, so too does the averaging period, which changes the interpretation of this average. There's a danger that if the averaging period is too long, the background concentration measurements may be representative of different synoptic periods.

There are definitely cases when a front is rolling through the area and the methane gradients are huge causing a background difference of 40 ppb or so.

Averaging over a long period of time smooths out these variations. This is why large sample size of estimations is suggested because unavoidably in any type of top-down experiment (aircraft or inversion) there are going to be days that are unsolvable because of complex background. The hope is that there are more days with homogenous background (background variability is less than city enhancement) than days with heterogeneous background (background variability is more than city enhancement). We can improve our chances by eliminating wind directions that are especially problematic.

Line 453-454: "Background random error is a function of sample size and decreases as a number of independent samples increase". As mentioned earlier, I don't think that this is the interesting part of the temporal variability analysis and is already a property of standard error. I think it would be more interesting to discuss how different the standard deviation in the background concentration differences are between wind directions. It would be interesting to know if the sample size at which the standard deviation of the background concentrations stabilizes is similar for all directions. This would be a helpful number if it can be assumed to be generalizable. Basing the minimum sample size on the standard error is less generalizable as the required precision may differ, and the variability in background concentrations would differ between regions.

Part of this question is already addressed above. It would be preferable to carry out an analysis presented in this paper for any other region in question where CH4 emissions are of interest. As you point out, each region is unique and presents its own challenges. But if one is able to understand what issues may arise when beginning their top-down estimation for a particular area, they may be able to avoid large errors simply by better constraining their experiment.

Figure 9: This caption should be expanded in order to make each figure stand-alone.

Done.

Response to Referee # 3

The narrative of the paper would be improved by more directly addressing the relationship between this paper and the previous investigations of methane from Indianapolis (all of the papers cited in Figure 1). The Introduction does a fine job of giving a general review of the previous studies, but the connection to the central problem of different studies/methods yielding different results is weaker in the rest of the sections.

We would like to thank the reviewer for their insightful suggestions. We tried to address noted issues to the best of our ability. Each comment of the reviewer is addressed below individually.

For example, how are the methods for background estimation in this paper different than the methods used in the previous studies?

This is a very good question. The description of methods to determine background used by other studies is now added to Section 2.4 (methodology section has been reordered due to requests from another reviewer).

The point of the article in question is to emphasize the challenge that background estimation may pose for flux estimation. There is no perfect method for background estimation; it is something that must always be addressed differently depending on a situation at hand. For instance, for aircraft mass balance studies there are 3 main methods to determine a background value. First is to pick a smallest edge value of a transect. Second is to linearly interpolate background field of a transect from one edge to another edge. Third is to use an upwind transect as a background field. In the case of an inversion, it is common to pick a tower that is located generally away from the sources and has on average smallest overall enhancement. In the current paper, background is chosen using 2 different criteria. For both of these criteria, each wind direction (using 8 main wind directions) is considered separately. This is possible because we have 9 towers and we can always change our background tower based on a wind direction. This is an advantage as with this strategy we may be able to better reduce contamination of local sources and to better represent upwind sources that are approaching the city.

In the Conclusions, you recommend a strategy for background estimation for the aircraft mass balance method, but you did not describe how it was done in the previous studies.

Here is what we wrote: "For the aircraft mass balance approach, we recommend an upwind transect be measured, lagged in time if possible, to provide a more complete understanding of the urban background conditions. Complex background conditions might suggest that data from certain days or wind directions should not be used for flux calculation." The strategy recommended is intended to better help an analyst to understand the background conditions for a given day. It may be possible that the day in question should not be used for flux estimation due to complex and variable background. We have not seen this type of critical discussion in the articles that employed aircraft mass balance methodology. Generally speaking, there are some "standard" methods of background calculation for an aircraft approach. Unfortunately no method is perfect and each one has its disadvantages. First is to pick a smallest edge value of a transect. This method could be wrong when the upwind plume is narrow and is not represented well by the edges of a transect. Second method is to linearly interpolate background field of a transect from one edge to another edge. This method is better, but it may not always account for the complex gradient of background that may occur over the plume. However, this method could help to identify that background is complex and the day should not be analyzed further for a flux value. Third is to use an upwind transect as a background field. This is potentially the best method if a case is in steady state, but realistically the issue here is lag. Plume is always moving, so upwind and downwind transects are not sampled simultaneously. If typical aircraft mass balance approach assumptions are satisfied this should work well, but from our experience that is not always the case and therefore a closer analysis must determine if a given day is acceptable for a flux estimation or not depending on how background is behaving.

So our goal was not to introduce a new background methodology for aircraft studies necessarily, rather to suggest caution when such data is analyzed. However, we did add the background estimation methodology of aircraft studies as you suggested (Section 2.4) as well as some recommendation regarding background for aircraft data analysis (conclusion).

Also, the Lamb paper identifies a major discrepancy between top-down and bottom-up estimation of the non-biological portion of Indianapolis methane *emissions, and the current paper is a follow-on to that paper, but it is not clear whether this paper resolves that question or not, or only partly resolves it.*

In this paper we are unable to address this question directly as we have no measurements of ethane (C2H6), which is a tracer gas used by Lamb et al., 2016 to separate biological CH4 from non-biological CH4. However, we can answer this question indirectly by estimating total emissions of the city and subtracting "known" biological sources (such as landfill, see the comment about landfill for more details) from that total. The residual is hypothesized to originate from non-biological sources such as NG. So we think it is likely there is no major discrepancy between top-down and bottom-up solutions. We think that some of the top-down solutions in Lamb et al. 2016 are biased high and should be lower more in line with bottom-up estimations. As we point out later uncertainty remains, but the high top-down estimates could be potentially explained by the erroneous assumptions in analyzes.

Specific Comments:

Line 19 – details about the type of analyzers and the measurement heights in the abstract are unnecessary and irrelevant.

The details are now removed.

Line 103: Please describe briefly how/why the landfill emissions are considered wellknown.

Originally this statement was based on the Greenhouse Reporting Data (GHGRP), which gave very similar emissions values for this landfill over 2010-2015 time frame. This is supported by Lamb et al. 2016 paper that also cited GHGRP as a reputable source for the landfill. However, 2016 and 2017 GHGRP indicates a 70% decrease in landfill emissions. That seems unrealistic given that our towers near the landfill do not show any decrease in average methane mixing ratios over these years in comparison with the previous years (Figure R1). We are currently in the process of investigating this discrepancy. So far we received no response from EPA regarding this.

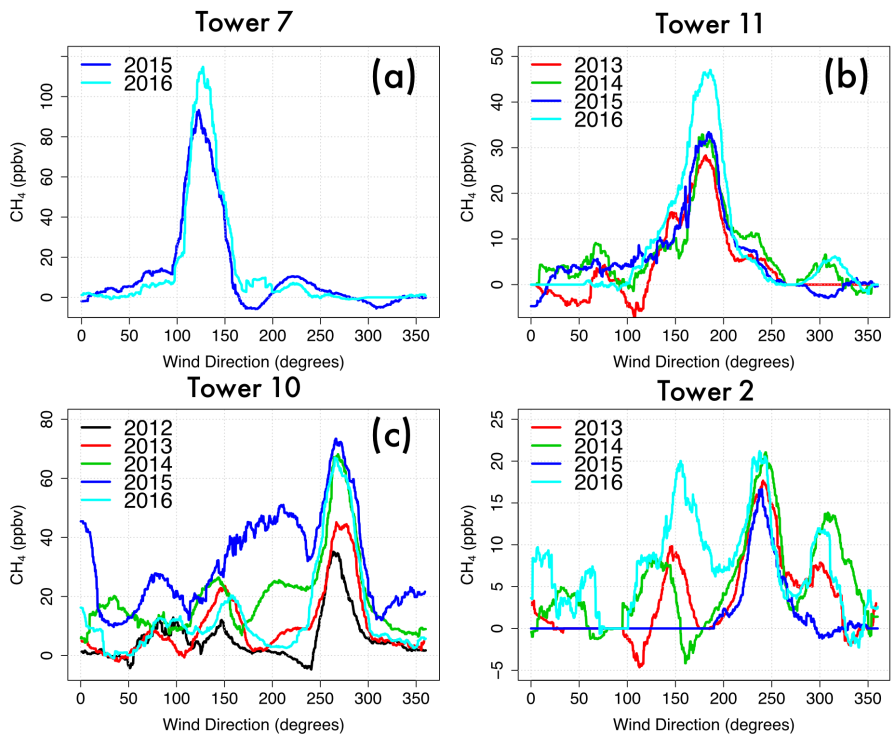

Figure R1. Yearly CH4 enhancement directional profiles for 4 INFLUX towers located in the city of Indianapolis: (a) tower 7, (b) tower 11, (c) tower 10, and (d) tower 2. Note that there is a lot more variability in the towers closer to the landfill (Towers 10 and 11). Often the plume is unable to fully mix in a close proximity to these towers resulting in the higher variability. Landfill peak is apparent in all of the towers. Tower two, which is more representative of a fully mixed plume, shows no dramatic change in the landfill emissions from year 2015 to year 2016.

For more details see (this needs to be copied and pasted into a browser): https://ghgdata.epa.gov/ghgp/service/facilityDetail/2017?id=1002683&ds=E&et=&popup=true

We do have some potential evidence for variability in landfill emissions from Cambaliza et al. 2015 article that used aircraft mass balance on five different occasions to calculate methane flux from this landfill. Their average is higher than GHGRP, but their estimation has high uncertainty. They were not able to make consistent calculations. Their city totals seem abnormally high on two of the flights, which may indicate there were issues with a background assumption. In addition, Cambliza et al. 2015 used California Landfill Methane Inventory Model v 5.4 or CALMIM to estimate methane emissions from South Side Landfill (SSLF), where results are more consistent with GHGRP.

Overall, we think we know an approximate value of the SSLF methane emissions, but this section in the paper about the landfill is now rewritten to reflect a truer state of knowledge regarding SSLF landfill emissions.

Section 2.2 – There are extra details here that are not very relevant to the current paper and have already been described in other papers. This section could be made more concise.

The section's length has been slightly reduced.

Section 2.4 (Sources) – Although references are given, the source of the numbers in this paragraph is not clear. Are they from direct measurements, an inventory, or something else?

The sources of theses numbers are a combination of bottom up and top down estimates to give a range of possible emission values. This section is now rewritten as another reviewer also asked about this.

Here it is:

2.7 CH_4 Sources

Only a few known CH_4 point sources exist within Indianapolis (Cambaliza et al., 2015, Lamb et al., 2016). The Southside Landfill (SSLF), located near the center of the city, is the largest point source in the city with emissions ranging between 28 mol/s (inventory) and 45 mol/s (aircraft) depending on an emission estimation methodology. SSLF could account for as little as 22% or as much as 63% of total Marion County CH_4 emissions (top-down from Cambaliza et al., 2015; invetnory from Maasakkers et al., 2016; inventory and top down from Lamb et al., 2016) contigent on how much of the total city emissions are coming from NG. Other city point sources are comparatively small; the wastewater treatment facility located near SSLF contributes about 3-7 mol/s (inventory from Lamb et al. 2016), and the transmission-distribution transfer station at Panhandle Eastern Pipeline (also known as a city gate and further in this study abbreviated as PEP) is estimated to be about 1 mol/s (inventory from Lamb et al. 2016). The remaining CH_4 sources, mainly from NG and livestock, are considered to be diffuse sources and are not well known. Potential sources of emissions related to NG activities include gas regulation meters, emissions from transmission and storage, and Compressed Natural Gas (CNG) fleets. These diffuse NG sources account for 21-69% (this value varies due to the uncertainty in SSLF emissions) of the city emissions or 20-64 mol/s (top down from Cambaliza et al., 2015; inventory from Maasakkers et al., 2016; inventory and top-down from Lamb et al., 2016). Livestock emissions for Marion County are estimated to be around 1.5 mol/s (inventory from Maasakkers et al., 2016). An important question remains of whether SSLF or NG is the dominant CH4 source in Indianapolis. There could also be a possibility of temporal variability in either of the sources as described in the section above.

Section 2.4 (Background) – As written, I had to read this section many times to try and understand it and I'm still not sure I fully understand the two methods, so it needs to be re-worked for clarity. Why is a viable method not to take the lowest measurement among all towers at a given hour as a background? How do these two approaches compare to those used in the cited aircraft and tower-based top-down studies?

We apologize for the confusion. Also this section should be numbered 2.5, we will correct the numbering in the next version of the paper.

The main point of this section is that we can pick multiple backgrounds for a given situation and it would be impossible to say which one is better. Hence, two arbitrary but acceptable backgrounds are chosen here to estimate methane enhancements. If background is uniform or closely so, as sometimes stated in literature, then we would see no significant difference between the enhancements calculated with different backgrounds. Yet we show here that this difference is significant and choice of background matters.

We edited this section to try to clarify this point.

It is not a viable method to take the lowest measurement among all towers at a given hour as a background because the background we are interested in is not always the lowest value as illustrated in the schematic shown in Figure R2. True background lies upwind of the city (or the downwind tower of interest from which the enhancement is calculated) and it is not the lowest value. Because we are trying to identify the enhancement specifically from the city we must subtract exactly what is coming into the city. The methane is heterogeneous as described earlier and therefore it is a challenge to identify exact background even at a not-so-large scale as Indianapolis.

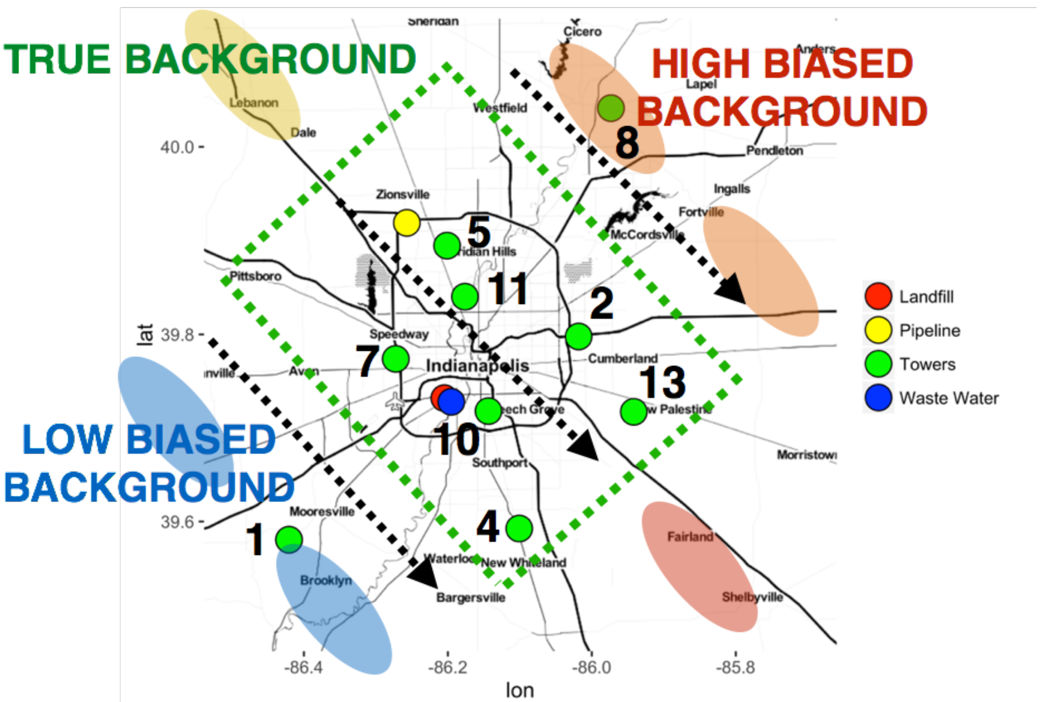

Figure R2. Theoretical representation of boundary layer CH_4 plume heterogeneity across Indianapolis and nearby areas when winds are from the northwestern direction. The colors indicate relative CH_4 concentrations where yellow is neutral, blue is low, and red is high. Green dashed lines indicate the assumed boundaries of Indianapolis. Also shown are INFLUX towers with CH_4 measurements and known sources.

The approaches used in Criteria 1 and 2 are not entirely different from aircraft and inversion studies, but it is exactly the point. All of these approaches are acceptable, it is just important to understand what kind of uncertainty they are causing. In some cases, we recommend not to perform flux calculation since the uncertainty is too large due to complex background. Identifying those complex days is whole another topic, which deserves a separate paper, but we do think this is very important and must be emphasized.

Here are a couple of figures that show the heterogeneity of methane in Indianapolis. Indianapolis CH4 observations indicate highly variable background with day-to-day variations at times reaching 150 ppb (Fig. R3). Similarly, WRF-CHEM simulations show occasional spatial non-uniformity of CH4 (Fig. R4).

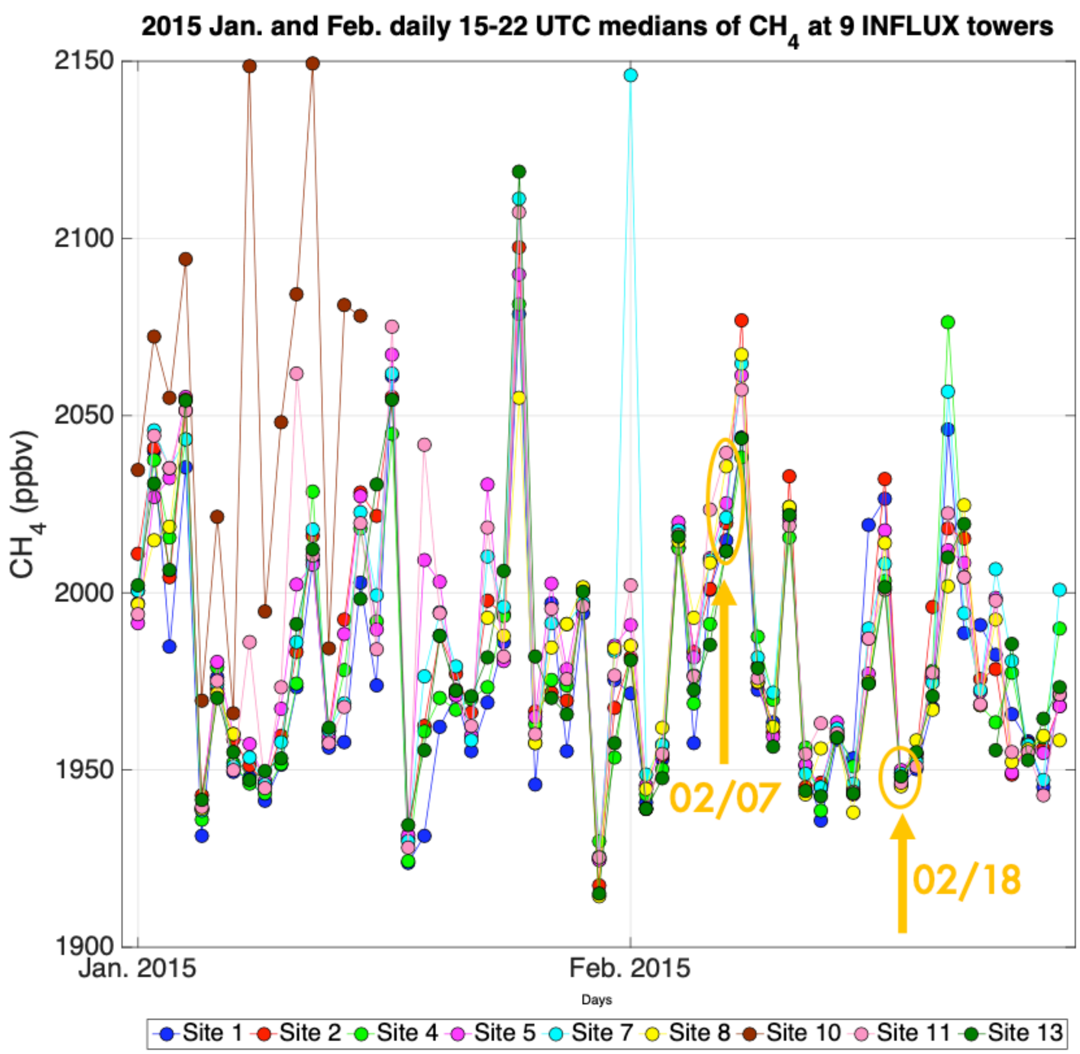

Figure R3. Daily CH_4 medians over 15-22 UTC at 9 INFLUX towers.

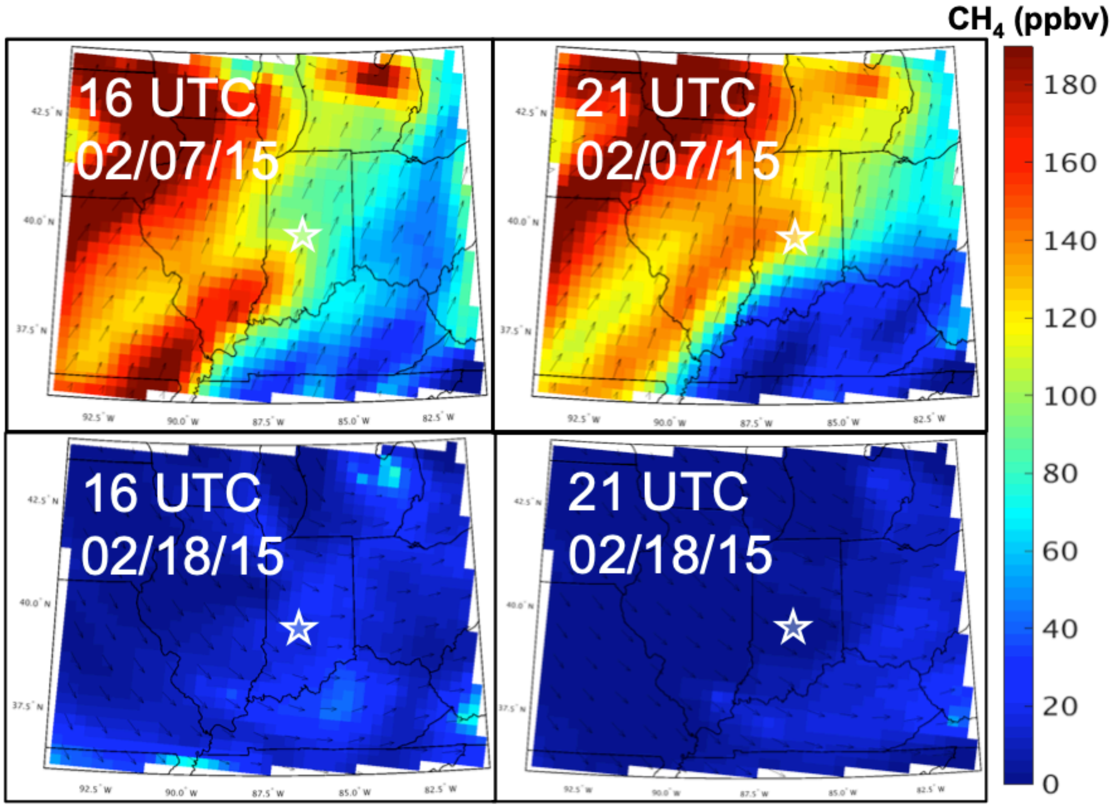

Figure R4. 27 km WRF-CHEM simulations of CH_4 enhancements (ppb) for 2 different days using EPA 2012 CH_4 emissions (Maasakkers et al, 2016): **(top left)** late morning on 02/07/15 **(top right)** late afternoon on 02/07/15 **(bottom left)** late morning on 02/18/15 **(bottom right)** late afternoon on 02/18/15.

Section 2.5 – How far away are your receptors and wind measurement locations since you say that this method requires them to be nearby?

In this case towers are the receptors and they are labeled as green diamonds in the figure below (Fig. R5). The wind data is gathered from the 3 stations identifiable by white colored shapes on the figure as explained in section 2.3. We found that the wind measurements are generally consistent between these 3 stations and therefore their combination is well representative of the city overall. Perhaps occasionally tower 8 may not be represented perfectly by these winds, but we do not think that such situation occurs often.

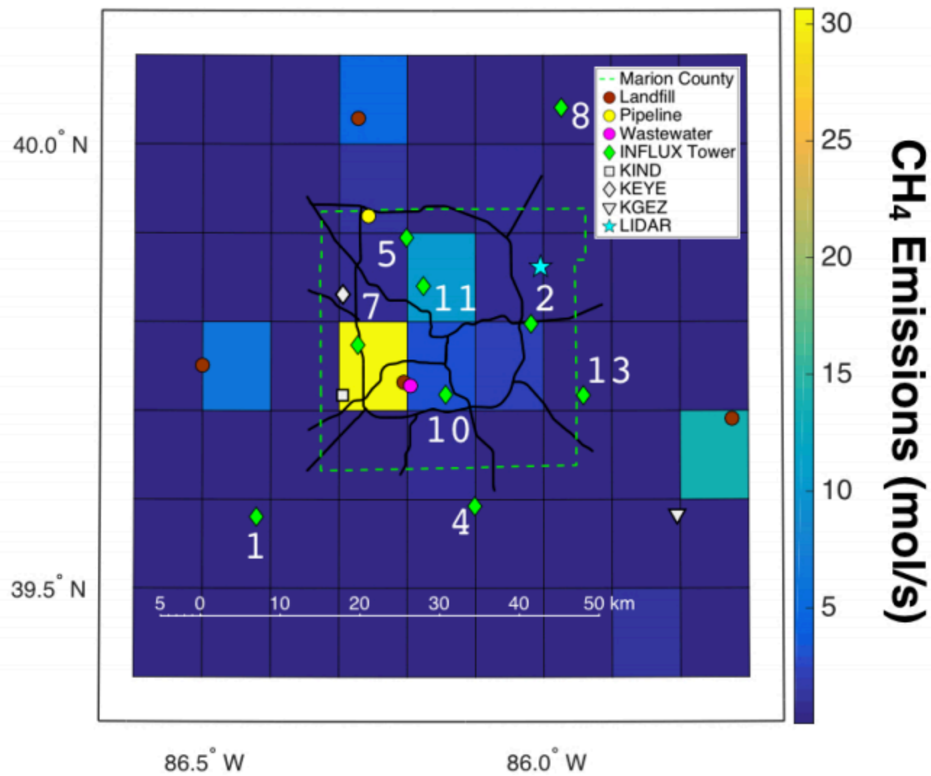

Figure R5. Map of the primary roads in Indianapolis, INFLUX towers, lidar system, weather stations, and a few CH_4 point sources plotted over the gridded CH_4 emissions (mol/s) from the EPA 2012 Inventory (Maasakkers et al., 2016). The gridded map of emissions includes emissions from these point sources; their position is provided to aid in interpretation of the observations. The dashed bright green line denotes Marion County borders.

Section 2.6, Line 261 – Which towers were used for which wind directions?

This question is answered at the end of Section 2.6 (lines 283-287). Here are the relevant sentences: "For S and SW wind directions tower 8 observations are used to represent downwind conditions with background observations coming from towers 1 and 13, respectively (based on Criterion 1 shown in Table 1). For W wind direction, tower 13 observations represent the downwind with background obtained from tower 1. The wind direction is required to be sustained for at least 2 hours, otherwise the data point is eliminated."

Section 3.1 - The first three sentences could probably be condensed into one concise sentence without losing any meaning. - I am having trouble squaring *your description of the domain differences with my understanding of the Lamb paper. By my reading, the Lamb paper describes developing an inventory for the larger domain, but you say that the inventory covered mostly only Marion County. I find it hard to believe that Lamb et al. would perform such a detailed analysis and accidentally compare totally different areas for the inventory and tower inversion. What am I misunderstanding? - With the revised inverse emissions estimate, it is not clear whether you've replicated the methods of the inversion in Lamb et al. over a smaller domain, or whether you've used the boundary layer budget method described in the method section.*

Unfortunately there is indeed an inconsistency between domains used by Lamb inventory and inversion. The domain used by the inversion contains 3 landfills that are not part of the Lamb inventory. We are not sure how this happened, but that is what we are seeing when we read Lamb's paper. The Figure R3 shows domain used by the inversion. You can find this domain in the supplemental of Lamb paper. We also have access to the prior used in the inversion of Lamb paper and these 3 landfills are in there. You can see landfills marked by the brown dot in the Figure R5.

The point we are trying to make here is that it is imperative to be very careful when such comparisons are performed. It may seem obvious that boundaries of emission areas need to be the same when they are compared, but it seems that occasionally such detail can get neglected when analysis considers many other complex parameters. In this we work we would like to emphasize the importance of this initial step.

The revised inversion estimate just shows what would happen if the original inversion had solved for Marion County. We asked the person who did the original INFLUX inversion to rerun his code for the whole region and then just for the Marion County. The result is shown in Figure 1 of the submitted article. Boundary layer budget method is not used for this result. It is used later in the paper to try to understand temporal variability of CH_4 flux in Indianapolis. Clarification is now added to section 2.1.

Section 3.2 - How much of the data are filtered using the criteria you give? - Line 372: Suggest: "Because Indianapolis is a relatively small emitter of methane, and because there are relatively large sources outside of the city, uncertainties due to background estimation are comparatively large."

The answer is at the beginning of the Section 3.2. Here are the relevant sentences: "To make the comparison as uniform as possible, only data from 12-16 LST are utilized (all hours are inclusive), when the boundary layer is typically well-mixed (Bakwin et al., 1998). A lag 1 autocorrelation is found between 12-16 LST hours, i.e., the hourly afternoon data are correlated to the next hour, but the correlation is not significant for samples separated by two hours or more. Therefore, hours 13 and 15 LST are eliminated to satisfy the independence assumption for hourly samples. Furthermore, we make an assumption that the data satisfy steady state conditions. If the difference between consecutive hourly wind directions exceeds 30 degrees or the difference between hours 16 and 12 LST exceeds 40 degrees, the day is eliminated. Days with average wind speeds below 2 m/s are also eliminated due to slow transport (the transit time from tower 1 to tower 8 is about 7 hours at a wind speed of 2 m/s)."

Because the city of Indianapolis is surrounded be sources that are similar to its CH4 flux magnitude it is not surprising that occasionally there are complex background scenarios that are difficult to address (include modeled map?). If Indianapolis was much larger source than its surrounding sources background would not be a big issue, but in this case it is not so. The goal of our background variability study is to show how variable CH4 background is on average at this location. The data was filtered only to eliminate extreme cases, but other cases, which are likely to be used by inversion studies, or even aircraft methodologies were left in. Another important point of our study is that inversions and mass balances should be carefully filtered to exclude complex background days. However, we did not see an evidence of that filtering in case of studies performed at Indianapolis.

Thank you. Suggestion is incorporated.

Section 3.4 - Isn't the result that the SSLF is the largest and strongest source in the city consistent with your prior understanding, as described in the methods? - Looking at figure 8, I can't tell which enhancements around T10 you think are from SSLF and which are from NG leaks. - How do you square your findings that emissions from NG is not a significant portion of emissions with the findings in Lamb et al. that approximately half the emissions are from NG using ethane as a tracer? Are you saying that you can't see the sources because they are below your detection threshold or that their signals or swamped by that of the SSLF, or are you saying that their existence is entirely not supported by the data? On the one hand, you *say there is not much evidence for a diffuse NG source, but on the other hand, even after adjusting the domain, your top-down estimate is still much higher than the inventories. You should at least acknowledge this contradiction or remaining possible existence of unknown sources. - Line 488: The description of "occasional" on seems incorrect since this apparent signal shows up in Figure 8, which represents a two-year average.*

It is indeed consistent with our prior understanding; however, there are views that natural gas could be larger than SSLF as a source of CH4. We agree that Figure 8 does not provide definitive answer, yet it allows us to see that the landfill is likely the strongest source in the city as no other point or area source is making such a noticeable enhancement at any of the towers. Landfill is located to the west of tower 10 and therefore the plume that arrives from the west is unquestionably belongs to landfill. Even in Lamb et al. 2016 we see similar result by observing low ethane to methane ratios from southwest of tower 11. In general, the ethane to methane ratios provide an insight into the composition of a given plume measured at a certain point, but overall city composition is hard to measure unless the tower is downwind of the whole city. The issue here is that plumes may come outside of the city and we have no way of separating plumes inside the city from the outside using tower 11. Lamb article does say that they ran an inversion of sorts, but unfortunately they poorly explain how they did it and the assumptions that were made in those calculations make the result largely uncertain. We spent sometime trying to understand what they mean by "source footprints" in their supplemental material (S3.4); however, this terminology is not standard and therefore we have no idea how they arrived at their answer. Also their sample size is just 11 days, which is not enough to make a definitive conclusion. And their domain includes sources outside of Marion County (they even mention multiple landfills), which makes this even more confusing.

With regard to adjusted inversion, there is still uncertainty in that result since it only used 3 towers at most (the tower record is sparse in 2012-2013 time frame). In our second attempt of running this inversion it seemed sensitive to the prior, which hinted that the system might need some more experimenting and testing. However, we admit that we cannot say for sure there is no significant diffuse source at Indianapolis comparable to landfill. But the evidence suggests that it is not as large as previously was suggested. The goal of this paper is to show that some of the very large values attained by the top-down estimates in this case appear to be unconvincing since adjustments in background and domain do affect the flux values. We are going to leave a space for potential diffuse source as suggested, but we must stress that its existence is highly uncertain given the data at this point.

Technical Comments

Line 61: Suggest: "...atmospheric methods and inventory assessment have sometimes succeeded..." Are there are cases when these two criteria have been met but reconciliation has not been achieved?

This depends on the definition of reconciliation. However, it may be possible that a study found an agreement between an inventory and a top-down methodology, while another study did not find that for the same region. Then perhaps reconciliation is under question.

Line 70: Suggest: "Recent studies of urban CH4 emissions in California indicate..."

Done.

Line 72: The phrasing "large NG infrastructures" is strange and evokes large individual pieces of equipment, which I don't think is your intent.

Done.

Line 79: Suggest: "in" → "for"

Done

Line 85: Suggest: ". . .comprised of irregular or periodic in situ aircraft measurements, continuous in situ observations. . ."

Done.

Line 91: Suggest: "well-suited" → "designed"

Done.

Line 94: Delete: "Recently"

Done.

Line 103: Suggest: "Uncertainty in total emissions is driven by. . ."

Done.

Line 132: Suggest: "tubes secured" → "air collected"

That is probably okay as it is.

Line 139: Suggest: "inflow" → "sample air"

Done.

Line 152: The given link re-directs to some other website.

Fixed.

Line 154: Suggest: "The anemometers are located at about 10 m AGL."

Done.

Line 200: Suggest: ". . .based on two different sets of criteria. Both approaches identify. . ."
Done.

Line 316: Suggest: "inventory" → "inventories (Fig. 1)"

Done.

Line 352: The meaning of the numbers "2 to 150" is unclear.

These numbers indicate an experiment sample size. Clarification is added. The idea is to see by how much the uncertainty decreases if the sample size is 150 (arbitrary large sample size) values. In theory each value could be used to solve for flux. But with sample size of 2 the uncertainty is large. This is an attempt to try to figure out how much data is optimally needed to solve for the emissions from the city. But due to various assumptions this is just an approximation. Topic related to this has been covered in great depth in the response to another reviewer. The revised article will contain some changes in that section.

Line 385: Suggest: “at least twice as high” → “approximately twice as large”

Done.

Line 396: Suggest: “did not change significantly between 2014 and 2016.”

Done

Line 519: “Dennis” – Do you mean Brian?

Yes, sorry.