Reviewer #1:

General comments:

1) They do not quantify mixed height in the downwind urban plume itself, so are blind to heat island effects that cause emissions to mix to a higher altitude (e.g., the Birmingham study described in Trainer et al. (1995)) than might be suggested by vertical profiles to the side of the urban plume;

This comment is perplexing as we did in fact quantify PBL heights through aircraft vertical profiles, and reported those data in the original submission. During the 01 June 2011 flight experiment, we actually conducted our second vertical profile within the boundary of the city, downwind of the city plume as shown in the original manuscript, Figure 1 (the first vertical profile was outside the city). The actual vertical profile data were both shown in Figure 2 of the original submission. The (single) vertical profiles flown on 01 March and 29 April 2011 flight experiments were also conducted downwind of the city. For the revision, we have added the flight paths for the 01 March and 29 April 2011 flight experiments in the supplementary information (Figure S4), to show where these vertical profiles were conducted. So, for each case, we did in fact fly vertical profiles downwind in the plume. We have further clarified this in the revised manuscript on page 12 lines 350 – 363. If there is an urban heat island effect, and we measure the height of the boundary layer downwind, then it has no impact on emission rate measurements, as discussed in the text. There is only a potential impact if the boundary layer height grows during the course of the measurement, resulting in entrainment of free tropospheric air, with impacts on the quantitation, as now discussed on pages 24 - 25 lines 729 - 747 of the revised manuscript.

2) They do not quantify detrainment (either episodic or continuous) of urban emissions from the boundary layer to the free troposphere, which can be exacerbated by elevated, buoyant plumes (e.g. the power plant CO_2 plume) and can further be exacerbated by the presence of an urban heat island, or in the presence of clouds;

We have now included a discussion of detrainment in page 21 lines 624-633 of the revised manuscript, which explains why we believe detrainment to be unimportant in our experiments. Our vertical profiles of CO_2 and CH_4 (now included in Figure S7 in the supplement) do not show any obvious evidence of measurable detrainment of CBL air into the free troposphere, most likely due to the close proximity of our downwind flight path to the city plume. Thus, we assume that detrainment is not important for our experiments.

3) They do not quantify uncertainties induced by non-constant wind field history, although they implicitly assume that winds measured during the downwind transects are unchanged since the time of emission. The flight patterns described in this study are not sufficient to quantify these effects (or demonstrate they were negligible) but these effects appear to not contribute to the uncertainty estimate. These must be treated rigorously in the uncertainty terms before the draft can be considered for publication. Including these and other terms in the uncertainty is a large reason the Trainer et al. mass flux for CO from Birmingham has a stated uncertainty of 100%. The present draft claims a significantly higher accuracy, which is unjustified until these additional sources or potential bias are addressed, and further is contradicted by the large spread in their own estimates, especially compared to Mays et al., to Hestia, and to the much more well-known emissions from the Harding Street power plant.

There are important differences between our approach and that described in Trainer et al. Among these is that we conducted a multi-transect approach to evaluate the mixing state of plumes in the boundary layer. With only one transect it is impossible to know if the boundary layer is well-mixed with respect to the local scale sources that are being quantified. On page 8, lines 234 – 238, we discuss more clearly that the experiment was designed to avoid accumulation. Specifically, the magnitude of the surface winds was monitored a couple of hours before the start of the experiment, to avoid conditions in which calm morning winds could lead to accumulation of emissions prior to outflow. In addition, we also added on page 23 lines 702 – 704 of the revision that the aircraft continuously measures the winds, and we use the measured changing mean winds in our analyses, effectively minimizing the uncertainty in this component relative to other studies. This also distinguishes our approach from that of Trainer et al. We also distinguish our results from those of Trainer et al. and all others by our experiments evaluating the precision, which is ±30% on average. This study is unique in this regard and these facts justify our lower uncertainty. In addition, the measurement uncertainty and precision of our CO_2 and CH_4 determinations are ±0.1%, certainly more than an order of magnitude better than for NO_x, which was quantified in Trainer et al. (though the uncertainty of that measurement was not specified in that paper). And, unlike the Trainer et al. case, in which the background was acquired from upwind transects, we derive the background from the edges of the plume that is being quantified, as discussed in the paper, and are thus not impacted by changes in the boundary layer height (for the background). It is thus expected for all these reasons that our uncertainty is much better than that of Trainer et al and others.

Specific comments:

1. Introduction

29898 line 25: for city-scale GHG estimates, should also cite Brioude et al. (2013) Top-down estimate of surface flux in the Los Angeles Basin using a mesoscale inverse modeling technique: assessing anthropogenic emissions of CO, NOx and CO2 and their impacts, Atmos. Chem. Phys., 13, 3661-3677, doi:10.5194/acp-13-3661-2013.

Brioude et al. (2013) has been cited in the paper. Thank you for the suggestion.

2. Methods

29900 line 25: "The city of Indianapolis: fossil fuel signal above background should be relatively easily identified." This assertion here in the text is premature; it was likely the basis for selecting Indianapolis as the focus of this study, but the qualitative assertion here is not supported by data at this point in the text. Can the authors provide some context? Signal above background implies some level of knowledge of background variability, so what is the enhancement above that variability expected from a US city of X million inhabitants into a mixing layer of Y meters at a wind speed of Z meters per second, etc. This is a non-trivial assertion. In the absence of large point sources (e.g., the Harding Street

Power Plant and Southside Landfill in Indy) the signal above background shown for Indy in Figs. 3 for CO2 and for CH4 appears to be quite small. Please rephrase here.

The assertion stated above that the Indianapolis fossil fuel signal above background is relatively easily identified is indeed observed to be case from data obtained by Mays et al. (2009). Using these data, we calculated the characteristic enhancement above the background in the urban air shed of Indianapolis, and found the result to be consistent with the tower measurements of Miles et al. (2013) as now discussed in page 5 lines 145 – 152 of the revision.

2.2 Aircraft-based measurements

29901 line 25 ff: "Ambient air : : : was pulled through 5 cm diameter PFA Teflon tube: : :" Teflon is not recommended for sampling CO_2 or CH_4. Why was Teflon used rather than stainless or Dekabon tubing, which are recommended and nearly universally used in the literature? Despite the short residence time in the Teflon, wall contact is the relevant parameter here – turbulence, esp. over the inlet entrance length, makes it probably that all of the sampled air sees an inlet surface. What effect could use of a non-recommended inlet material have on the stated accuracy for CO_2 and for CH_4? The discussion of comparability between CRDS and flask sample data on page 29903 does not quantify potential bias due to use of Teflon for sampling lines upstream of these two detection methods.

We now state in the revised manuscript, page 6 lines 180 - 181, that Teflon was used to minimize adsorptive losses for volatile organic compounds (VOCs), which are also sampled with the aircraft. As stated on page 6 lines 177 - 179, with a blower speed of 1840 L min^{-1}, the residence time in the manifold is 0.1s, which allows for the efficient transfer of gas from the nose of the aircraft to the point of sampling.

2.3. Experimental Flight Design

29903 line 21: ": : :flight experiments were conducted between 11:00 LT and 16:00 LT when the boundary layer was essentially fully developed." This seems not to be the

case for the flight of 1 June 2011 when the boundary layer grew by nearly 50% over the duration of the airborne study. Please rephrase, and define LT as local time please.

We repaired the text on page 8, line 230, and we defined LT as local time on page 8, line 229.

29903 line 28: ": : :transects : : : were conducted downwind: : : to the top of the convective boundary layer." A crucial but neglected assumption is buried here. The authors do not discuss the potential for detrainment of urban emissions from the convective (boundary) layer into the overlying free troposphere as a potential source of bias in their calculations of the urban flux. By neglecting to sample systematically above the boundary layer at the downwind transect location, they have no ability to detect or quantify detrainment of the plume. This potential loss process adds bias by removing plume mass from the boundary layer. Assuming no wind shear (in either velocity or direction) between boundary layer and free troposphere, a crosswind pass just above the boundary layer at the downwind transect location is key in detecting the presence of detrainment of urban plume mass. Given any wind shear between mixed layer and free troposphere, however, it becomes more difficult to detect (much less quantify) the potential for detrainment, as the plume aloft can experience very different transport than the plume remaining within the boundary layer. This additional source of uncertainty is not discussed in their error analysis and could potentially add significantly to the overall uncertainty. This process may have contributed to the large day-to-day variability observed for the Indy urban plume in CO_2 and in CH_4 in this work and the Mays et al. It is difficult to assess quantitatively the degree to which detrainment might have affected the results; however, it is a crucial, potentially large, and certainly variable term in the uncertainty that appears to be wholly neglected in this report.

As stated above, and as discussed and presented in the original manuscript, we did sample above the boundary layer in every experiment. The text provided on page 9 lines 255 – 257 and on page 12 lines 353 - 354, and in Figures 1 and S4 (supplementary information), presents and discusses the vertical profile measurements in the downwind plume. For our relatively close downwind distances, detrainment is observed to be not important as shown in the vertical profiles of CO_2 and CH_4 in Figure S7, and we now state this on page 21 lines 624 - 633 of the revision.

29904 line 12: Another key, but unexamined, assumption is embedded here. Calculated boundary layer heights from a single location outside of the urban plume (two locations on 1 June) are used to estimate boundary layer height. The presence of any urban heat island effect could bias the flux calculation low, as the urban plume would be mixed to higher altitudes than suggested by the vertical profiles outside of the urban plume. This issue was identified in the Trainer et al. reference for the Birmingham urban plume, in large part driving the much larger uncertainty derived by Trainer et al. for the flux of CO from that city. To be fairly assessed for Indy, this report must additionally consider the effects of urban heat island driving a potential bias in downwind mixing heights (or leading to detrainment from the boundary layer over the city) that are not captured by the flight patterns in this study.

While we did not discuss the urban heat island effect in the original manuscript it was indeed effectively assessed in the measurements of the boundary layer height in the downwind plumes. We did not ever estimate the boundary layer heights, they were always measured. We clarified in the methods section (page 9 lines 255 – 257 and on page 12 lines 353 - 354, and in Figures 1 and S4 (supplementary information) of the revision that our vertical profiles for all flight experiments were indeed conducted downwind of the urban plume and thus, properly accounting for the potential urban heat island enhancement in the CBL depth downwind of the city. We added figures (Figure S4) in the supplementary information to show the flight paths for the 01 March and 29 April 2011 flight experiments, and where the vertical profiles were conducted.

29905 Equation 1: The limits defining the vertical column over which the data are integrated are given as z=0 to z=zi. Here z=0 is ambiguous – on page 29912 the lower limit

to z is described as the ground height, but in Eqn. 1 z=0 could be interpreted as mean sea level. Suggest explicitly setting the limits as z=z(sfc) to z=zi in Eqn. 1. What limits were used for the calculations in the paper? If the surface height of Indy was used, great, otherwise using z=0 meters will introduce an error. Please clarify in Eqn. 1 and verify that the integrals have been calculated between surface height and zi, not sea level and zi.

We have clarified on page 9, lines 272 - 274, and in Eqn. 1, that the integration starts at the physical surface.

29905 line 10: with transects at multiple altitudes at a single downwind location, the authors have chosen to interpolate in the vertical. This may be problematic (despite its use in Mays et al.) and certainly introduces non-physical features in the resulting 2D curtain plot. One, emissions from a non-buoyant source (such as landfill CH4 emissions) are mixed by turbulent transport to the aircraft transect altitude. It is unlikely that enhancements observed aloft are actually disconnected from the ground – i.e., the kriging interpolation routine results in a non-physical lofted plume, and suggest a minimum in CH4 below the minimum altitude of the aircraft transect, and other minima between each aircraft transect altitude. These are likely artifacts of the interpolation routine. While buoyant or lofted plumes such as the Harding Street power plant CO2 plume (Fig. 4) can mix downwards from enhancements injected above the surface, these features should not persist downwind over multiple convective cycles in a well-mixed boundary layer, and likely should not be observed in Fig. 4.

Two, the enhancements observed at different crosswind locations at different transect altitudes, which lead to variability in the interpolation, are just as likely due to wind speed or direction differences at the time of emission (remember, transects at different altitudes are separated by 20 minutes or so) as they are due to incomplete vertical mixing. I'm left wondering what quantitative value does this interpolation add to the data? It does produce a striking visual but the statistics are no longer robust in the interpolated field. The text must do a more thorough job justifying the quantitative use of interpolated fields, and correct the non-physical interpolation artefacts between transect altitudes and below the

minimum flight altitude, or simply derive their fluxes using the measured data and not the interpolated fields.

Many previous cases in the literature use one transect and assume complete mixing, without any knowledge as to whether or not this is in fact the case. So, one thing that differentiates our approach is that indeed we have multiple transects, and they generally show that unless you are at least 40 km downwind from point sources, the plumes will not be well-mixed through the boundary layer, and so single transect approaches produce results that are inferior to our approach, as stated in Section 3.5 (pages 30 – 31, lines 896 – 942) and Table 4. The discussion regarding Table 4 indeed indicates that our results and approach are improvements over the single transect, full-mixing assumption approach, which in many cases is simply not correct. Having more information is always better than having less. So, the variability from spatially variable plumes is measured in part in our precision measurements, and we have now expanded the discussion along these lines on page 25 lines 757 – 764 as well as on page 28 lines 845 - 860 of the revision. The reviewer is correct that the lack of data between the lowest transect height and the surface is an issue that was not fully discussed in the original version. We have thus added a component to the uncertainty analysis, from upper and lower limits for the values interpolated to the surface. We discussed this in detail on page 14 lines 407 – 415, and on page 23 lines 684 – 696 of the revision. Thank you.

29905 line 24: ": : : the section in the transect outside the projected city limits is used to calculate the mean background concentration: : :" This neglects a third major assumption: it appears that upwind variability is assumed to be zero for the purposes of the calculation, but this assumption is not stated and its uncertainty is not included in the error estimate. Later in the report the use of two aircraft is mentioned to permit a more thorough assessment of upwind variability, but it cannot be neglected here. Imagine a hypothetical upwind plume of a few ppm in CO2 (equal to the variability in the background boundary layer) underlying part of the Indy plume. The background assessment from outside the projected city limits would not capture that, and its contribution would unfairly be ascribed to Indy sources. This is typically a negligible bias for point source plumes, characterized by extremely large enhancements compared to ambient variability, but it

cannot be neglected when assessing the relatively smaller enhancements that make up a significant portion of the enhancement from an urban area. Since this cannot be corrected for by only flying downwind, it must at least be included in the uncertainty estimate, which appears to have been neglected in this report.

The reviewer is correct that this is an issue. We have found that the results are compromised by assuming that an upwind transect provides the best background, because, assuming that it is flown at an appropriate time (e.g. representing the transit time for air passing over the city) before the full set of downwind transects, there is normally a change in the boundary layer height, which changes the background concentrations. The edges of the downwind transects are preferred because they are obtained at the same time. We now clarify these on page 32 lines 957 - 968 of the revision. We further state that our revised procedures are now that we fly an upwind transect specifically to identify point sources flowing into the city. This is now explained on page 24 lines 724 – 729 and on page 32 lines 958 - 969 of the revision.

29907 throughout: Here the authors discuss boundary layer growth in good detail. However, this report appears to neglect an additional error term I was expecting to see here – how is detrainment, or mixing of urban CO_2 and CH_4 from the boundary layer into the free troposphere – handled? The lack of a horizontal transect above the mixing layer is a critical omission here. The Trainer et al. paper cited for the Birmingham plume analysis found a significant amount of urban emissions above the mixed layer due to detrainment. Here, vertical profiles outside of the downwind urban plume are unable to assess the extent to which detrainment may or may not have affected the observed Indy BL enhancements downwind. Since detrainment can be significant and episodic, and can be further exacerbated by urban heat island effects (see Trainer) its neglect in this report needs to be corrected. I suspect the measurements are not sufficient to assess this term quantitatively, but it must be included in the uncertainty estimate. Its contribution can range from negligible to significant, and will result in a low bias if not accounted for. I further suspect this contributes to the large variability in derived fluxes for CO2 and CH4 between transects and between flight days in this report.

As discussed above, we do have the information, and routinely fly a downwind vertical profile, as originally shown by our flight paths in Figure 1 and Figure S4 (supplementary information). These downwind profiles (Figure S7) have never revealed evidence for measurable detrainment, presumably because of the proximity to the sources and the time scale for mixing to the top of the BL and detrainment, as we now discuss on page 21 lines 624 - 633, where we indicate that we have made this assumption based on our observations.

29912 line 3: here the lower limit of column density is clearly indicated to be surface height, not mean sea level. What is the average surface height for Indy? What value was actually used in the calculations?

We indicate the average surface height for Indianapolis on page 5 lines 141 – 142 of the revision. All integration was carried out from the surface level (ground level) to the top of the boundary layer as described on page 9 lines 272 – 274.

29914 line 5: It appears, given the invariant values of the backgrounds (392.6 ± 0.5 ppm CO_2 and 1880.6 ± 2.6 ppb CH_4) assumed on 01 June, that the signal from Indy is dominated by the point source emissions from the Harding Street power plant (for CO_2) and the Southside Landfill (for CH_4).

We quantified this, and discussed it in detail in the original manuscript for CO_2. In the revised manuscript, this information can be found on page 29 lines 881 - 883 and in Table 1. The HSPP contributes about one half of the total CO_2 emission of the city, determined from three afternoon flight experiments in 2011. The Southside landfill represents about a third of the CH_4 emission, though that is the subject of a separate paper.

Their contributions are quantified later in the text, but it is difficult to visualize the signal from the on-road mobile sources of CO_2 given the time series presented in Figure 3. Please consider modifying Figure 3 to give more space to the key chemical parameters (perhaps shrink or remove the H_2O and altitude panels) and include lines and shading to indicate background values of 392.6 ± 0.5 and 1880.6 ± 2.6 in the time series panels of Figure 3.

This would illustrate graphically the magnitudes of the enhancements from non-point sources of CO2 and CH4 to the total signal from Indy, and indicate the sensitivity of the flux calculation to a non-zero upwind variability in the background. Put another way – how does the ± 0.5 ppm uncertainty compare to the actual enhancement shown in Fig. 3 for CO_2 (and similarly for CH_4)?

We accordingly modified Figure 3 to show that the urban plume is defined above the variability in the background. Thank you for the suggestion.

29914 line 14: Turnbull et al. (2013) is listed as "in preparation", but is not available so difficult to judge appropriateness here. What is ACP policy on this kind of citation? Not sure if that's a robust reference at this point.

We rephrased page 19 line 563 to indicate that Turnbull et al. is currently an unpublished work that is in preparation.

29916 line 14: Cambaliza et al. (2013) is also listed as "in preparation" – not sure if that's a robust reference, and it is not included in the references list at the end.

We are about to submit this work. It has been included in the reference list. Thank you.

29918 line 2: ": : : the history of horizontal winds prior to the experiment can also be important in the mass balance." Yet another critical point. The assumption is not stated clearly, but all the calculations assume that the winds measured during the horizontal transects are the same as the winds at the time of emission for the sampled air parcels. This assumption is very clearly spelled out in White et al. and in Trainer etal., but seems to be glossed over here. Any systematic change in wind speed between time of emission to time of measurement results in a direct bias (low or high, depending) in the flux calculation. Assuming measured winds from the aircraft are the correct value to use introduces additional uncertainty in the derived flux, and again I don't see this explicitly included in the uncertainty analysis. The authors must revisit the error analysis to include the several

additional sources of error identified in this review before this draft is ready for publication.

Please see the response in no. 3 in the General comments above. In addition, just as in Trainer et al. (1995) and White et al. (1983), we assume that the winds at the time of measurement are the same as at the time of emission, and this is stated in page 12 lines 347 – 349.

29918 line 24: having two aircraft, one upwind and one downwind, would help reduce errors in assumptions inherent in the flux calculation for area sources. But to realize the improvements described in this draft, the two aircraft would need to be exercised in a purely Lagrangian fashion, so that the same air parcels sampled on the upwind transect by aircraft 1 are sampled again on the downwind leg by aircraft 2. In practice this is difficult, and mixing creates additional hurdles to meeting the Lagrangian criterion in any case.

Yes, we agree, it is difficult because of the boundary layer height changes. The first aircraft is most useful for identifying upwind plumes, as we now state in the conclusions, and for greater sampling density in the downwind plume.

Alternatively, rather than 2 aircraft, the authors neglect to mention that a single aircraft with longer endurance, better speed, and increased range (relative to the one used in this study) can perform both upwind and downwind sampling, thus affording the same advantage as would two aircraft.

We thank the reviewer for these helpful points. We have included these points in our Conclusions.

29920 line 2: Please confirm the dairy cattle population data are relevant to the flight dates in question. Are these numbers averaged over any specific period of time? Some dairy farms can have very high fluctuations in their populations, which can add episodic variability to their emissions.

The dairy cattle population data from the Indiana Department of Environmental Management are the relevant ones for the 2012 flight date considered in the analyses.

29921 line 19: mixing height is assumed to be equal to cloud base. Here again the potential for detrainment, or venting into the free troposphere, appears to be unquantified by the aircraft data and neglected in the uncertainty estimate. This should be corrected, especially given the Harding Street power plant CO2 emissions are released at stack height and likely as a buoyant plume (thus more likely to be vented than a non-buoyant plume released at the surface).

Given that our flight tracks were considerably close to the source (5, 8 and 17 km downwind), the potential for detrainment is minimized. While meteorological conditions prevented us from performing a vertical profile on this flight date (01 June 2012), we have shown from our vertical profiles from previous experiments that there was no measurable evidence of detrainment (Figure S7, supplementary information). We note that these vertical profiles were about 40 km downwind from the city center.

29923 line 4: "However, the observed CO2 flux: : : was 60% smaller than reported by EPA." Again, is detrainment or venting of the elevated, buoyant power plant plume a source of this discrepancy? This complication needs to be addressed in the uncertainty if not in the calculation itself. Further, in the recommendations section at the end, it would be appropriate to call for another horizontal transect above the mixed layer height to account for this issue in future Indy experiments.

As we have explained in our original submission (page 29 lines 876 - 879), the presence of low level cloud layer (cloud base was 640 m) prevented us from probing the upper section of the boundary layer (due to FAA visual flight rules) leading to an expected underestimation of the emission rate. And, as explained above, we always conduct a vertical profile in the downwind plume.

Conclusions

The conclusions paragraph should be rewritten to reflect the expanded uncertainty budget, including the several additional uncertainty terms identified in specific comments above. The final paragraph seems highly speculative, especially in regard to an airborne mass balance flux experiment for a geographically larger source in complex terrain such as Los Angeles. Given the complexities of recirculation, stagnation, orographic lifting, and venting through multiple mountain passes (and indeed directly up mountain slopes) that are well documented in the Los Angeles basin, the simplistic flow-through model upon which INFLUX and this report are based on is likely inappropriate to apply to Los Angeles quantification. Recommend removing this paragraph – its assertions on the tractability of Los Angeles for this type of study are not supported by the known complexities of LA transport, from lidar studies in the 1980s through the NASA and NOAA airborne measurements in 2008 and 2010.

We have modified the conclusions section, and removed the paragraph pertaining to the application of the approach to Los Angeles, and we agree with the comment. We thank the reviewer for the helpful suggestion.