Anonymous Referee #2
Received and published: 18 January 2016
C11683 ACPD 15, C11683–C11687, 2016

The authors present an analysis of CO2 and CH4 fluxes using three years of data collected from the CARVE tower in interior Alaska. The manuscript is well-written and figures are easy to read. The measurements are of high-quality and made using established methods. There are some methodological issues with the analysis that need to be addressed but if these comments can be resolved, the paper should be published in ACP.

Major comments:

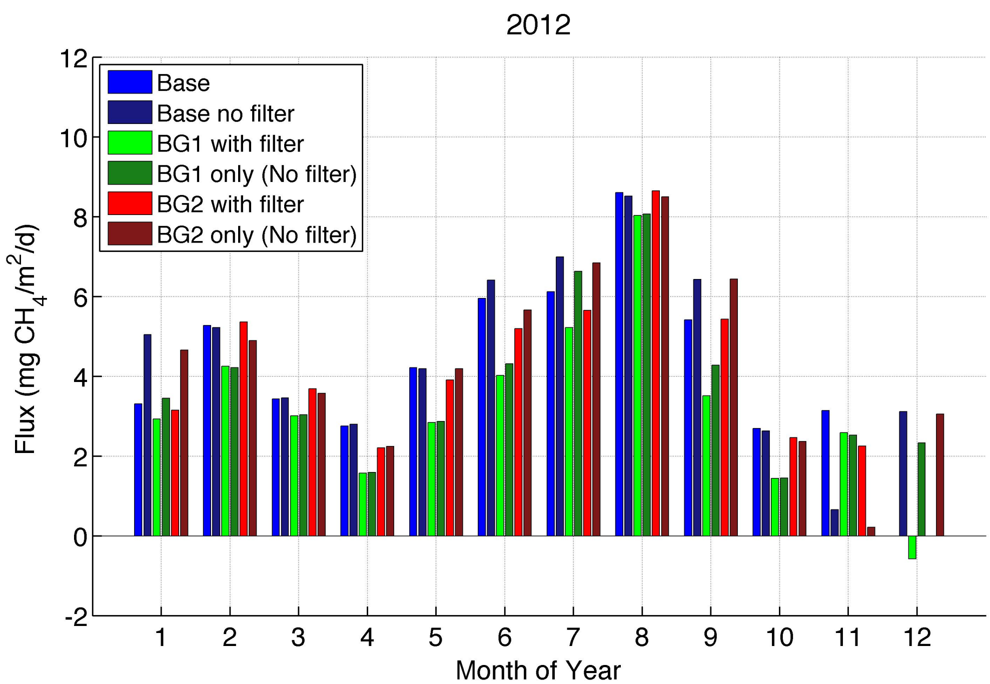
1. *The method used for the background identification has some issues. Firstly, the authors use a Pacific curtain as the background for calculating enhancements in the measurements. This only applies to air masses that are transported from the West. The authors first discard the subset of particles that did not exit from the West or exited from too far North-West and allow up to 25% of particles to be discarded in this way. Can it be clarified that both the effect on the surface influence as well as on the background are removed in this way? This would lead to up to a 25% error in the analysis, as of course the measurements would still be sensitive to these directions. Secondly, as the authors correctly discuss that the choice of background is critical (especially for CH4), what is the potential impact of these choices on the results? Can a sensitivity analysis be done? Can the authors use a better method for identifying the background, in the absence of having an additional tower that samples the background (i.e. using model simulations for other directions)?*

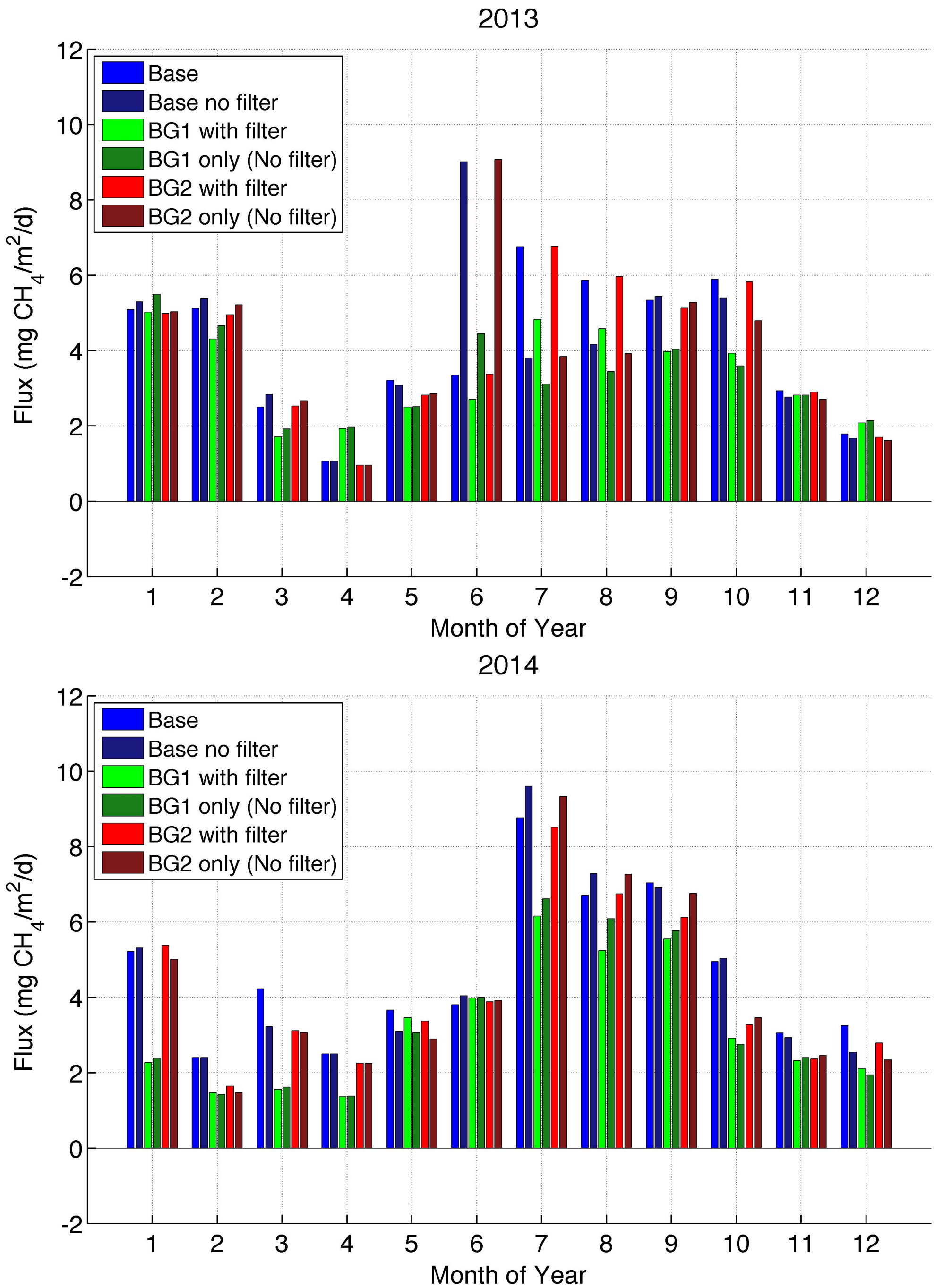
We now clarify in the text that only the background calculation is affected by discarding 25% of the particles. This has no effect on the footprint itself in our analysis, only in the background mole fraction assigned to that footprint. The reviewer is correct in then noting that hours with < 25% of particles discarded would have an additional error in the analysis, although the magnitude is not 25% - it is the difference between what the real background concentration would be if those 25% of particles were included (and their initial concentrations were known) and the background concentration we calculate using the other 75% of particles. As multiple reviewers suggested, we tested the effect on our final flux estimates by calculating two additional possible backgrounds as a sensitivity analysis:

a) BG1: Removing the background filter entirely: retaining all particles and if they do not exit to the West then tagging them with the background curtain mole fraction at their latitude, altitude, and time at their origin 10 days prior to the observation. No hours were removed from the analysis due to air masses beginning in the East.
b) BG2: Retaining particles that do not exit to the West in the calculation, but still eliminating the hourly observation from the analysis entirely if less than

75% of particles exit east of 160W and below 4000 masl (same filter as original setup).

To further examine our filtering choices, we also examined the effect of the choice of eliminating hours with high variability, large vertical gradients, or CO enhancements. Each of the above background (BG) scenarios was propagated to a flux estimate calculated both including and not including this additional filter. Six scenarios were investigated in total and the results for each year shown below in the figures. The bars for the "base case" (blue) represent the final flux values shown in the manuscript. The green bars represent background condition (a) above (BG1), and the red represent the results for background condition (b) (BG2). For each color, the darker shade represents the results if the additional filter is not applied. Thus, the dark green bars represent the unique case where there is no added filter at all – all existing observations are included and there is no filter for either the particles not entering from the West or for any variability, lack of vertical mixing, or biomass burning/pollution.





Here we also report the effect of both the background choice and the filter on the number of observations retained in the annual analysis.

Over all 3 years, 50% of hours were removed due to the combination of all filters; most were eliminated due to the background restriction. We understand that these are large numbers of points to be filtered, but we have chosen to filter them out because a proper calculation cannot be made with those observations.

Our analysis of these alternative scenarios showed that the CH4 fluxes from the different scenarios could be higher or lower than our base case, with the mean difference over all months and scenarios of 9-20%. Unlike the other scenarios, the case of BG1 (green), where there was no removal of hourly observations or particles based on their exit longitude, almost always resulted in lower CH_4 fluxes (9%-30% lower depending on the year). We found that the footprints whose particles do not enter from the West within 10 days of the observations are stronger, meaning the observations have a larger surface influence. Enhancements in CH_4 are not significantly greater, so that the flux estimate needed to achieve those enhancements is smaller. However, without a proper background calculation during these times, it is difficult to say whether the associated enhancements are correct. The differences remain within our 1-sigma estimate of background error. The text in Section 3.3 and 3.5 addresses this sensitivity analysis now.

2. In the absence of being able to do a full quantitative study with an inverse modeling analysis and appropriate treatment of boundary conditions, the authors need to more explicitly that they state that the quantitative results will be subject to larger uncertainties and biases (the qualitative results presented have more weight than the fluxes and uncertainties presented).

We agree, and have now added sentences in Sections 4.6 in the text emphasizing that our uncertainty analysis only covers one of many components of the total uncertainty on the flux. We have also emphasized in our abstract and conclusion the qualitative findings of the analysis of the tower data: the small signal, the importance of background in light of this small signal, and the general agreement of a simple CO2 flux model with the observations.

3. *The prior flux maps for CH4 need more discussion. Why is a uniform map assumed? What sources is this based on?*

The spatial distribution of CH_4 sources in Alaska is not well-known. Although maps of wetland extent (and other CH_4 flux maps) do exist, they vary, and, previous studies have shown little or no correlation with observations of fluxes (Chang et al., 2015). Evaluating different models and performing an optimization was outside the scope of this work. Instead, we opted for a simple analysis to interpret the CH_4 enhancements at the tower. A follow-on study is in preparation (Miller, S., A. Michalak, R. Commane, C. E. Miller, J. B. Miller, R. Chang, S. Dinardo, J. Henderson, A. Karion, J. Lindaas, and C. Sweeney, "A multi-year estimate of methane fluxes in Alaska from CARVE atmospheric observations") that focuses on methane only and uses additional observations to perform a geostatistical inverse analysis. We now discuss our choice further in Section 3.5.

4. The method on how CH4 fluxes were estimated is not clear. Was this not done for CO2 because it was shown the correlation between modelled and observed CO2 enhancement lay close to the 1:1 line (i.e. fluxes were correct?). This is a bit confusing because several times, it has also been stated that CO2 signals are larger than predicted. In Sec 3.5, the CH4 flux estimation method is described. Observed CH4 enhancements are first averaged into daily values and then into monthly values. The modelled CH4 enhancements were first simulated on hourly timescale but then averaged into daily and then monthly. Why is the flux estimation based on monthly values rather than averaging higher resolution fluxes? How is this fit done (there is no information on what method is used to determine the optimal fit)? Further in section 4.5, the fitted monthly fluxes are then used to drive daily enhancement values using the WRF-STILT footprints. The correlation analysis is then based on observed versus modelled daily enhancements (modelled values based on the fluxes derived from the monthly fit). In effect, the authors are fitting fluxes based on the monthly enhancement values, but then state that the daily modelled enhancements do not match the observations well. Can the authors firstly reorganise these two sections together and secondly, explain why the different timescales are used? What is the implication that fluxes derived based on a fit to monthly data does not lead to a good fit on daily timescales? Why is hourly not used (as was done for the CO2 correlation analysis)? This section needs some more explanation.

We have now added equations to the text in section 3.5 explaining the CH4 flux estimation method. Scaling fluxes to the monthly mean is equivalent to taking the average of the daily fluxes, but weighting each by the footprint influence on that day (similar to Chang et al, for their vertical profile analysis). The equations should explain the fit as well – there is no fit performed, only a simple multiplicative scaling on a monthly basis. In section 4.5, we show how the results compare with observed higher resolution observations. We have changed these to hourly in the revised paper (from daily), to be consistent with the CO2 analysis, as this and other reviewers suggested. We have chosen to retain the organization, keeping the methods and the results in separate sections, with references to each in the text.

5. The flux uncertainties that are presented are likely to be too small. The authors are only including an uncertainty due the background term. In almost all inverse modelling analyses, studies have shown that the overwhelming uncertainties are due to model, representation and aggregation errors. In addition, uncertainties should include uncertainties in the fit. While the authors do mention the simple uncertainty analysis that was done, it should be more explicitly stated that the uncertainties presented are likely to be an underestimate of the true uncertainties for these reasons.

We have now added more explicit language and new sentences in the text in section 4.6 that the error bars on the CH4 flux estimates represent only one of many uncertainty components, and are an underestimate of the true total uncertainties.

Minor comments: Abstract lines 19-22: These two sentences seem inconsistent. Firstly that the model does well in predicting magnitudes and distributions, and then that the signals are larger than predicted.

We agree this was confusing, and have added text to the abstract to indicate that although the model performed very well in general, the exception to this was the underestimate of late fall respiration.

P. 34874 line 4: Typo 'in' P.34876 line 16: Are these supplemental flasks to the in situ? Flasks are not discussed until later.

We fixed the typo and removed the reference to the flasks in this section, so they are only explained later on to remove confusion on this.

P. 34876 line 23: How can the authors be sure that there is no local influence at the site? A tower at 32m could still very much sample local fluxes. More discussion is needed on this.

We have removed the claim that the observations are less likely to be influenced by local sources, because we have not expressly shown this is the case here. We remove the local source influences as much as possible in our analysis by filtering the data for vertical gradients and high variability.

P. 34877 line 11: What are the three heights? Not mentioned until later.

We have included the three heights here now.

P. 34878 line 6: How long is the flushing time?

Three minutes of data is discarded for flushing, although this is conservative for the higher-flow analyzer.

P. 34878 line 9: How often is the water correction performed? Is it instrument specific?

It is instrument-specific (we have now added this information in this section), and was conducted prior to deployment for each. Because we swapped out the units several times, it was not done on a regular schedule, but the timing of the deployments is outlined earlier in this section.

P. 34878 line 17: How do the authors know that the nonlinearity has not changed during the four year period? Has it only been measured once prior to deployment?

We have added text in this section to further explain our calibration method. Two tanks deployed on site have varying mole fractions for each gas, so that a change in the slope or non-linearity would be detected from their residuals.

P. 34880 line 11: What is the size of the domain?

The domain for the WRF-STILT model is 30N to 90N and 180W to 180E (i.e. the entire globe in longitude). This is now inserted in the text.

P. 34880 line 13: Has any sensitivity analysis been done to show that 500 particles is sufficient to statistically represent the footprint? If 1000 particles were used, what is the effect? This is an issue with LPDMs in general that is never adequately discussed.

STILT is computationally more intensive than Flexpart or Hysplit, which usually are run with a larger number of particles. STILT has been run with 500 particles in this and other studies, and for the model runs presented here, there was no sensitivity analysis on the particle number performed. There have been two previous studies that have shown that 500 is generally a sufficient number of particles for this kind of application. Hegarty et al. (2013) tested STILT with 500 and with 5000 particles, and found little difference between the two cases in replicating a dispersion experiment:
Hegarty, J. D., R. R. Draxler, A. F. Stein, J. Brioude, M. Mountain, J. Eluszkiewicz, T. Nehrkorn, F. Ngan, and A. E. Andrews (2013), Evaluation of Lagrangian Particle Dispersion Models with measurements from controlled tracer releases, *J. Appl. Meteor. Clim., 52,* 2623-2637.

In addition, Gerbig et al. (2003) (see section 3.1) also performed a sensitivity test using STILT with different numbers of particles between 50 and 1000 to estimate the error due to the stochastic nature of the model; at 100 particles it was estimated to be 13% and random (i.e. with no correlation between different receptors).

Gerbig, C., J. C. Lin, S. C. Wofsy, B. C. Daube, A. E. Andrews, B. B. Stephens, P. S. Bakwin, and C. A. Grainger, Toward constraining regional-scale fluxes of CO2 with atmospheric observations over a continent: 2. Analysis of COBRA data using a receptor-oriented framework, J. Geophys. Res., 108(D24), 4757, doi:10.1029/2003JD003770, 2003.

P. 34880 line 17: This could lead to important differences. Can the authors do a sensitivity test to show the effect of using model ground level?

We have now performed a sensitivity test to quantify the effect of running the model from different altitudes on the CH4 flux estimate. We analyzed the difference in footprint influence and how it would affect our monthly flux estimates. We found minimal effect between runs at 35, 100, and 300 magl for the filtered observations during the months of March-September of all three years (5-9%). However, in winter months, the fluxes were decreased by 12-17% when using the 100 magl footprints and by 26-32% using the 35magl footprints. We have added a paragraph in Section 3.2 detailing these results. Although this sensitivity analysis points to larger uncertainties in transport during the winter months, we still believe that using the higher altitude is will be more accurate than the lower altitude, given the limits of the model to appropriately account for the ridge-top location of our site. We now add the mean summer and winter differences for each year between flux estimates calculated at 35 magl and 300 agl to our uncertainty analysis.

P. 34882 line 4: To clarify, are these particles removed from the footprint calculation as well? As discussed above in the major comments, this artificially changes the footprint. For example if a plume were travelling to the south-west corner, such that 75% exit "west" and 25% exit "south", by removing the southerly particles, the plume will have narrowed by 25%, which is significant.

We do not remove the particles from the footprint calculation. The footprint is intact and used as long as at least 75% of the particles entered from the west, only the background assignment is based on fewer particles. (Now clarified in Section 3.3.) However, as we now show in Figure 1, the average footprint is changed by the elimination of footprints where > 25% of particles enter from the East.

P. 34883 line 17: What about the lowest height on the tower? Is it used for anything?
Added sentence in section 3.1 that the lowest level is not used.

P. 34885 line 1: How is this scaling performed? What statistics?
We have clarified this method here by including equations describing the method more fully. The scaling factor was equal to the ratio of observed to modeled monthly CH_4 enhancements, and was different for each month.

P. 34885 line 3: See major comment about uncertainties and being more explicit that these uncertainties are likely too small.

See response to #5 in major comments.

P. 34886 line 7: As the authors state, the magnitudes of amplitudes are difficult to compare as they are very site dependent and depend on sampling height, PBL height etc. This section doesn't add a lot because of all of these additional complexities and any comparison is quite speculative. The section would be better if shortened and condensed to say that the range of measured fluxes is xx and that this is dependent on factors such as: : :

We have shortened this text.

P. 34890 line 20: Following major comment above, this section on the correlation analysis of CH_4 enhancements is confusing.

We have added some text to clarify this section.

P. 34891 line 4: This could also be due to uncertainties in the background estimation
Yes, this is added here now.