

# ***Interactive comment on “Estimating regional scale methane flux and budgets using CARVE aircraft measurements over Alaska” by Sean Hartery et al.***

**Sean Hartery et al.**

rachel.yw.chang@gmail.com

Received and published: 1 July 2017

## **1 Summary**

**This paper uses aircraft observations of in situ trace gas concentrations and thermodynamics to constrain Lagrangian-transport-inversions of CH<sub>4</sub> flux during a campaign over the Alaskan region as part of the CARVE project. More specifically, spatially resolved fluxes of biogenic CH<sub>4</sub> flux during the growing seasons between 2012-2014 have been derived. The work uses a rich measurement dataset that has been carefully obtained and calibrated to a high standard.**

**After a description of the measured dataset, the work goes on to describe a flux**

Printer-friendly version

Discussion paper



derivation method using footprint sensitivities and inversions calculated using WRF-STILT. These methods are interesting and adapt existing conventional optimal flux transport inversion approaches to attempt to link spatially-resolved flux to soil temperature and depth as a function of time. Such an attempt is highly challenging and this paper is a trailblazer in terms of attempting to do this from a long-term aircraft campaign.

That said, I do have some major questions and concerns about the methods and the conclusions drawn from them. At the moment, the flux work seems to be predicated on assumption after assumption, followed by a procedure of arbitrarily discarding data (over half of it in the end), temporal and spatial averaging, and averaging some more, and then discarding more data, before using only 68 (or is it 146?) measured aircraft profiles (averaged to mixed-layer partial columns) to obtain biogenic growing season flux as monthly-averages over 3 years. This equates to about 4.5 profiles per month (though this is only a rough average calculation as there is no information on the sampling statistics per month other than the total number of profiles used across the whole study). I find it hard to accept that such a limited dataset can derive robust flux statistics representative of regional monthly means, especially where the only uncertainty given on the fluxes is the standard error on the mean of the already monthly-averaged fluxes. Such an error statistic is useless – it neither represents the systematic error associated with the method, nor represents the natural variability of CH<sub>4</sub> flux in the region between each independent flux calculation. Instead, it convolves the two with no possible determination of which dominates.

What I instead believe the authors have here are a set of independent flux retrievals (one per profile/flight) and independent posterior flux uncertainties using their method. The extensive averaging of these independent retrievals in the paper make it very difficult to assess the performance of the method; and the current error budget is meaningless. Until I can see more about the performance

[Printer-friendly version](#)[Discussion paper](#)

of the method and the statistics of flux retrieval-by-retrieval, I don't have any confidence in the conclusions and discussion later on (e.g. on soil temperature relationships).

The paper does present some very interesting analysis and I believe there is some really great science to come from the work. Therefore, I definitely recommend publication in ACP as the methane flux problem is a key topic in atmospheric and geoscience at the moment and this paper represents an exciting way to make use of long-term aircraft datasets. However, I do have to recommend major revisions at present as I think the way the analysis has been done needs to be extensively rewired to present more meaningful data that the reader can more transparently assess, especially with regard to independent flux calculations and uncertainty and error budgets. I'll try to give some specific constructive guidance on this below, which I hope would help will turn a questionable analysis into something really quite interesting and useful. My review won't discuss the flux-soil relationships as until I can see the results from the revisions below I don't feel I have enough information to assess the later aspects of the paper.

We thank the reviewer for their very thorough reading of the paper and constructive comments and suggestions. We believe that the assumptions made in our analysis are not as extreme as the reviewer suggests and that our responses below address some of the confusion caused by our initial descriptions. We have added extra text in the manuscript to clarify these points. We agree that quantifying the uncertainties is very challenging. Since our analysis method is not a Bayesian inversion, rather we simply divide our observed column methane enhancements by the modelled column footprint sensitivity to estimate net fluxes, the model does not actually provide a posterior flux uncertainty. However, we now explore the uncertainties associated with our analysis through bootstrapping / Monte Carlo. This is explained in more detail below and is

[Printer-friendly version](#)[Discussion paper](#)

described in a new section (Sect. 3.7) of the manuscript.

To address the reviewer's suggestions, we have added a Supplement which includes tables listing all the profiles, whether they were included in our analysis, and the reason for exclusion if relevant, as well as monthly means and uncertainties calculated using different methods. It also includes additional figures showing the flux calculated from individual profiles with uncertainties as described above, the resulting residual when compared to the monthly means and background CH<sub>4</sub> observed from the aircraft, at the BRW observatory and inferred from the CRV tower.

### Specific comments

**1. Abstract, line 9 etc: It is very important to note that the whole analysis of the paper derives “net emissions”, not “emissions”, e.g. the statement that “. . .Boreal emissions. . .accounted for the remainder of the emissions” should contain the word “net” as the study does not address local or regional sinks (albeit potentially small). This important point needs to be kept in mind throughout the paper when discussing flux and needs to be very clear to the reader early on.**

The reviewer brings up a very important point and we have changed the wording throughout the text to reflect both that we are only calculating net emissions and that we are assuming that the emissions are originating from biogenic sources.

**2. P.2. line 31-32: I agree that scaling local fluxes to regions is challenging, even in areas where it may be argued it is possible to derive meaningful regional statistical parameterizations (such as this paper sets out to do). However, there**

[Printer-friendly version](#)

[Discussion paper](#)



are some studies (not currently cited) that have attempted to do this (also at high latitudes) using a combination of aircraft, chamber and eddy covariance measurements. It would be useful to discuss and cite such work in this paper (as it seems very relevant to the introduction, and later discussion, in this paper). Please see: O'Shea, S. J et al.: Methane and carbon dioxide fluxes and their regional scalability for the European Arctic wetlands during the MAMM project in summer 2012, Atmos. Chem. Phys., 14, 13159-13174, doi:10.5194/acp-14-13159-2014, 2014.

A citation to this study has now been included.

### 3. P.5: Footprint method:

The authors have used WRF-STILT to derive a surface sensitivity footprint. The footprint seems to have been derived using a grid with frequency/counts equal to the summed incidences of residence of 500 reverse- Lagrangian particles per measurement in the “lower half” of the boundary layer. This is a can of worms and it is glossed over far too quickly here, and later on. I guess this definition is a bit arbitrary and I have plenty of sympathy with it, as defining surface influence in Lagrangian trajectories is very difficult to quantify. However, I would have liked to have seen more discussion on the uncertainty or sensitivity that this arbitrary definition of surface contact may have on the derived flux footprint. The authors at the very least need to clearly acknowledge that there may be an unquantified source of model transport error coupled with the use of their lower half PBL definition; and - more usefully - they could examine footprint sensitivity to different surface contact definitions (e.g. as other percentages of the boundary layer height). I raise this simply because there is no basis to believe that the lower half of the PBL is in dynamical contact with the surface, especially in enormously diurnally-variant boundary layers such as those in Arctic spring.

[Printer-friendly version](#)[Discussion paper](#)

It could be argued that the diurnal ventilation and contraction (i.e. entrainment and detrainment) of the boundary layer could skew the footprint derived here to one more biased toward representing daytime flux (as daytime PBL trajectories that are isentropically detrained into a PBL-residual layer at night-time would not be counted at night-time in the footprint using the authors' method along the 5-day history used). This could have implications for the fluxes that are derived and their biogenic interpretation and quantification. I have no major problem with the use of an arbitrary definition such as this, as it attempts to do the best it can with the information available, but I do think the reader needs to be made more aware of the potential limitations and issues with it.

And some of this may be quantifiable with a sensitivity analysis to PBL depth-contact versus footprint. Without this, I would have some outstanding questions about the numerical validity of the later flux calculations and what they truly represent. I liked the discussion on what the footprints represent more globally on Page 6 but more needs to be added. In summary, I suggest an easy (making it clear as to the limitations) fix and a bigger effort (sensitivity) fix to help those following the work to make their own informed judgment. Page 9 lines 1-5 seem to suggest that some effort has been made to examine footprint sensitivity to the inclusion (or not) of discarded profiles so perhaps it could be simple and useful to add some of this to the paper to convince the reader that there is no important bias (even if as an Appendix?).

The reviewer brings up some very important concerns that face any receptor-oriented atmospheric model. The initial description in the manuscript was rather sparse so we hope that the discussion in response to this and the next two points will provide more clarity. The text in Sects. 3.1 and 3.4 of the main text have been edited to include more details.

In an early publication describing the development of STILT, Gerbig et al. (2003) in-

[Printer-friendly version](#)[Discussion paper](#)

investigated the effects of varying the assumed height at which the air is well-mixed. In their sensitivity analysis, they found no significant changes in their results when the fraction was varied from 10–100% of the boundary layer, although lower fractions resulted in less particles influenced by the surface and therefore increased noise in their results. We acknowledge that their findings were for continental North America using assimilated meteorological data, and may not be representative of Alaska, especially in the spring. Nevertheless, these findings suggest that this parameter is not the greatest source of uncertainty in our analysis. A sentence in the main text now directs the reader to the Supplement where this discussion is now included.

Air detrained into the residual layer at night is not influenced by the surface and should therefore not be included in the footprint sensitivity. As the 500 particles are traced backwards from the receptor point, a number will be in the PBL during the day. As detrainment occurs at night, some fraction of those particles will move into the residual layer while the rest remain in the nocturnal boundary layer. Of the particles remaining in the boundary layer, those that are in the lower half are the ones actually influenced by the surface and therefore included in the footprint sensitivity in the evening hours. We acknowledge, however, that defining the boundary layer at night is difficult and now mention it in this section. Nevertheless, we have fairly high confidence in the representativeness of our footprint sensitivities since past CARVE studies using the same WRF-STILT model coupled to a surface CO<sub>2</sub> vegetative flux model have successfully predicted atmospheric CO<sub>2</sub> mixing ratios in the mixed layer (Commane et al. 2017, Karion et al. 2016). Since CO<sub>2</sub> fluxes are bidirectional and highly diurnal, these studies suggest that the transport model represents surface influences during both day and night reasonably well. This is now discussed in Sect. 3.1 and reference is made to these other studies.

**4. Related to the comment above, were 500 particles released per singular Picarro measurement, or were 500 particle released per mixed-layer column**

[Printer-friendly version](#)[Discussion paper](#)

**used in the later inversion? If the latter, were these equally spaced with altitude in the mixed layer? A sentence to make this clear would help.**

As described in the text, the Picarro measurements below 1 km were aggregated into discrete bins every 5 km in the horizontal and 50 m in the vertical and above 1 km they were aggregated every 5 km in the horizontal and 100 m in the vertical. This resulted in approximately 23 000, 32 000 and 36 000 receptor points in 2012–2014, respectively. At each receptor point, 500 particles were released and traced backwards in time for five days. At each hour backwards in time, a surface sensitivity map is generated that represents the surface that influenced the measurement from that hour. These 120 surface sensitivity maps (24 h  $\times$  5 d) are added together to determine the cumulative 5-day footprint for a given receptor point. As described on P 8, lines 4–9 of the original submitted manuscript, the total footprint sensitivity associated with a vertical profile was determined by further binning these receptor points every 250 m and calculating the mean footprint at each height. The resulting profile of footprint sensitivities was then summed to determine the total surface influence on a given profile. The text in Sect. 3.4 has been improved to make this more clear.

**5. P.6. Section 3.3 - Mixed layers: This relates very closely to the comment above. If mixed layer height (i.e. instantaneous local PBL height plus residual layer height) is used to derive the fluxes later on, how do you differentiate between the true PBL (which is in contact with the local surface) and the residual layer (which may represent the previous day's local PBL ventilation, or advection from a more distant regional source)? This is very confusing. The method states (page 6, line 26) that the authors use the integrated mixed-layer partial column concentration (not the vertically-resolved measurements) to calculate flux in Section 3.4. This must then surely convolve any local emissions (in the true local PBL) and non-local emission (in any residual layer). The authors later go**

[Printer-friendly version](#)[Discussion paper](#)



to great lengths to show that any residual layer is not influenced by long-range (non-regional) transport but this does not solve the problem of varying air mass histories for the true PBL versus the residual layer when these get clumped into a partial column for the purposes of the inversion. When this singular column concentration is used in STILT and coupled to the footprint described in Section 3.1 (and the issues alluded to in the previous comment), it seems impossible to deconvolve spatially-resolved flux with any true or traceable footprint sensitivity as the column represents an unknown mix of local and non-local surface contact. Again, I have a lot of sympathy (more than it may sound like) with the approach and doing the best job possible with adjoint models. But there is currently no awareness or clarity of these issues in the text which would alert the reader to the challenges and limitations in the approach.

Page 7 goes on to demonstrate that 50 percent of profiles were discarded from flux analysis because mixed layers could not be reliably separated or that PBL was “over-estimated” when compared to trace gas profiles. What does over-estimated mean here, and why are the remaining profiles more trustworthy for analysis than those retained? Discarding half of the dataset to select only those data that agree with a contrived and questionable model is a little worrying. I would have liked to know more as to why >50 percent of the dataset is at odds with the model/assumptions and to know what the sensitivity to including the data might be. Can you be confident that the way the remaining 50 percent has been retained hasn’t led to some systematic bias in the data and model treatment?

We agree that the partial column is influenced by both local sources in the boundary layer and regional sources in the mixed layer. The important point to note is that the spatial distribution of the footprint sensitivities of air sampled in the boundary and mixed layers will reflect local and regional sources, respectively, thus correctly attributing sur-

[Printer-friendly version](#)[Discussion paper](#)

face influences to the source regions.

The 273 profiles quoted in the main text represents every instance of the aircraft flying from the surface to >3 km or vice versa. The reviewer will appreciate that not every aircraft transit between these heights properly sampled the entire mixed layer and sufficient height into the free troposphere to properly determine the background CH<sub>4</sub> mixing ratio. In addition, there are times when the atmospheric structure does not follow the structure of a classic textbook profile, with mixed layer heights from observations of water vapour, potential temperature and other trace gasses differing, either with each other or with the refractivity method. The cause of these inconsistencies would warrant a study of its own in boundary layer meteorology and is beyond the scope of our study. We therefore limit our calculations to profiles that are well-defined and where we believe we understand the dynamics.

As requested by the reviewer, we now include calculated net fluxes for all identified profiles in the Supplement. However, we do not believe that all of these calculated fluxes are representative of actual net surface emissions and therefore do not include flagged profiles in any of our calculations. The effect of excluding these profiles from the monthly mean cannot be determined. It should be noted that for all 273 profiles, the mean 95% confidence interval (C.I.) for an individual CH<sub>4</sub> flux estimate was 8.1 mg m<sup>-2</sup> d<sup>-1</sup>, ranging from 0.54–34 mg m<sup>-2</sup> d<sup>-1</sup> (2.5%–97.5% percentile). However, for the subset of 146 profiles used in our analysis, the average 95% C.I. was 5.2 mg m<sup>-2</sup> d<sup>-1</sup> and ranged from 0.6–15 mg m<sup>-2</sup> d<sup>-1</sup>. As stated at the end of Sect. 3.4, we can say that based on the model footprint sensitivities, our sampling region is not skewed by excluding these profiles.

**6. P.8. Using CO as a tracer for combustion CH<sub>4</sub> sources: What about pure fugitive emissions of thermogenic CH<sub>4</sub> (where there is no combustion)? This could conceivably lead to an over-estimate of biogenic flux if the remaining profiles contain any significant non-biogenic CH<sub>4</sub> from sources not co-emitted**

[Printer-friendly version](#)[Discussion paper](#)

with (potentially large fluxes of) CO. I see that only 9 of the profiles were discarded by this definition and the analysis of sensitivity by including them to derive a different flux is useful. This style of analysis starts to give the reader what they need to assess things.

As the reviewer pointed out in the first specific comment, our calculations are only representative of net methane emissions in our study region, to which we attribute to biogenic emissions. These assumptions and other possible sources of methane are now discussed in Sects. 3.4 and 3.5. It should be noted that preliminary results by Flerchinger et al. (2016) on the North Slope of Alaska suggests that fugitive emissions from thermogenic sources are small compared to the net emissions and mostly confined to the region near Prudhoe Bay. Although these results are still preliminary, they indicate that our attribution of emissions to biogenic sources is reasonable.

**With this in mind, I have a few suggestions below to consider:**

**Suggested principal corrections:**

**7. The monthly-averages given may be hiding a wealth of useful data. A time series of biogenic flux (as an area-normalised quantity – i.e. as biogenic flux per unit time per unit area) for each independent flux retrieval (I believe there are 68 of these?) would be useful. The posterior flux uncertainty (that STILT should yield as output) could be plotted as an error bar on each data point on such a plot.**

As discussed at the end of Sect. 3.4, 146 profiles are used to calculate the monthly means. As requested by the reviewer, we now include a time series of the flux calculated from each individual profile (Fig. S1) as well as the residuals from the

[Printer-friendly version](#)[Discussion paper](#)

monthly mean (Fig. S2) in the Supplement. Our inversion method is based on simple linear regression and not a Bayesian analysis, therefore there is no posterior flux uncertainty from the model. Instead, we derive our uncertainties in the estimated flux by bootstrapping each element that goes into the calculation (observed mixing ratio, pressure, temperature and footprint sensitivity at each height bin, as well as background methane mixing ratio and mixed layer height). Over 500 iterations, we arrived at 95% confidence intervals which are shown as the uncertainties in Fig. S1. A detailed description of this analysis is now included in Sect. 3.7 in the revised manuscript.

**8. Error/uncertainty analysis: As discussed above, the current tolerance placed on the derived fluxes is meaningless and does not represent either systematic error (flux inversion uncertainty) or natural variability. The seasonal trend plotted in Figures 4 and 5 do not convince me that natural regional variability dominates the mean as this could simply be a manifestation of the changing northern hemispheric seasonal background and priors used. I would suggest that the posterior flux uncertainty of independent retrievals/footprints is used instead as this captures the uncertainty on each retrieval. And then, rather than a standard error on the mean flux (taken from the spread of the averaged inverted fluxes), which would clearly be an incorrect (and much reduced) error, I would recommend quoting the posterior flux uncertainty (calculated as an average of the posterior uncertainties across all inversion that contribute to the final monthly mean). It would be important to give the average of the posterior uncertainties (not their standard deviation or standard error), to yield a meaningful uncertainty on the monthly mean flux. Such an error will still convolve natural flux variability but at least it would be a more accurate measure of the systematic uncertainty in the method used. This should replace the shading (error bars) used in Figure 4.**

[Printer-friendly version](#)[Discussion paper](#)

We thank the reviewer for this useful suggestion. We averaged the 95% confidence intervals as the uncertainty in an individual flux estimate to determine the uncertainty in the monthly mean as well as calculating a footprint weighted average. These results can now be seen for each month and on the growing season budget in Tables S5–S7. We chose instead to represent uncertainty in the monthly mean with the standard deviation weighted by the inverse of the 95% confidence intervals because they tended to result in larger uncertainties in the budget and we felt represented the variability associated with each month's measurements as well as the uncertainty associated with individual flux estimates. We now discuss these findings in the newly added Sect. 3.7 and have replaced the shading on Fig. 4 to reflect these new calculations.

**9 A table of the derived mean fluxes (and their corrected uncertainties) could be presented, which displays flux with and without the sensitivities/assumptions that have been used (i.e. masking seas and mountains, removing elevated CO profiles). These fluxes are currently in the body of the text, making it hard to compare them. And perhaps a new flux could be calculated where all 248 profiles are used in the inversion? A table would add (at a glance) the comparison between these sensitivities. This latter sensitivity test would then give the reader all the information they need to compare the information and make their own judgment about what they trust and the implications of the assumptions used.**

Tables S5–S7 in the Supplement now show the monthly mean net fluxes. As we discuss in the response to point 5 above, we do not believe that the excluded profiles provide meaningful estimates of net fluxes, and therefore do not include them in any monthly mean calculations. However, at the reviewer's suggestion, we now include flux estimates for all the individual profiles so that their quality can be judged by the reader.

[Printer-friendly version](#)[Discussion paper](#)

**Technical corrections:**

**10. 1/ Title – hyphenate “regional-scale”**

**11. 2/ Abstract line 1: change to “. . .gas but its emissions. . .” Not “their”.**

**12. 3/ P.1. Line 10: change to “. . .CH<sub>4</sub> flux was. . .” or “. . .CH<sub>4</sub> fluxes were. . .” - There seems to be some confusion between the use of singular and plural references throughout when referring to flux and fluxes, respectively. Please check as I won't list any further instances.**

**13. 4/ P.2 line 5: change to “40°N” and check throughout. “40N” is not acceptable.**

**14. P.2. Line 9: century should always be capitalized when referring to a specific century.**

**15. P.4. Line 6: add space to “195 K”.**

**16. P. 8, line 11: typo - change to “weighting”.**

These have been corrected in the revised manuscript.

**References**

Floerchinger, C. R., McKain, K., Newberger, T., Handley, P., Wofsy, S. C., Sweeney, C., and Benmergui, J. S.: A methane budget from the North Slope of the Alaskan Brooks Range

[Printer-friendly version](#)[Discussion paper](#)

including arctic tundra and proximate oil and gas operations, American Geophysical Union, San Francisco, CA, 12-16 December 2016, 2016.

Gerbig, C., Lin, J. C., Wofsy, S. C., Daube, B. C., Andrews, A. E., Stephens, B. B., Bakwin, P. S., and Grainger, C. A.: Toward constraining regional-scale fluxes of CO<sub>2</sub> with atmospheric observations over a continent: 2. Analysis of COBRA data using a receptor-oriented framework, *J. Geophys. Res.*, 108, n/a–n/a, doi:10.1029/2003JD003770, <http://dx.doi.org/10.1029/2003JD003770>, 4757, 2003.

---

Interactive comment on *Atmos. Chem. Phys. Discuss.*, <https://doi.org/10.5194/acp-2017-72>, 2017.

[Printer-friendly version](#)[Discussion paper](#)