

Review of Erland et al., Comparing Airborne Algorithms for Greenhouse Gas Flux Measurements over the Alberta Oil Sands.

This is fundamentally a good solid piece of work which is worthy of publication, it is very useful to have data compared in this manner and builds confidence in the methods being applied to convert measurements to fluxes. It is frustrating that activity data isn't available (and I'm sure the authors are more frustrated than me by this) as it would make conclusions and discussion so much stronger. I have a number of minor comments / requests / clarifications which I would like the authors to consider.

L19. Is surveys a better term than samples? Samples gives the impression of very limited data collection.

Abstract general – it would be good to highlight how the lack of on-site activity data can clearly have a profound impact on the depth of conclusions and understanding to such work given the variability of emissions over several days.

L53 (and in other places too). Can the references be split out from the long list so that they match with the parts of the list where they apply?

L69 Feels repetitive from previous paragraph – suggesting rewriting / removing first part.

L80 Replace “Recently, ” for “Here,” or “In this work,”

L82 un-needed “and”?

L99. Clarification needed here. Coincident sampling stated, but then alludes to periods where sampling conditions are the same... If coincident then should be identical? Maybe needs clarification over what time-period different methods require and therefore over what period conditions need to be stable to give comparable results.

L102. How much is expected to be under reported? A numerical range or approximation would be useful for scale of potential problem.

L106. This is a long sentence and could be reframed within a specific “aims” paragraph? It would be good to mention how this type of work fits with program initiatives such as OGMP v2.0 to encourage operators to properly measure emissions rather than estimating from emission factors / engineering calculations.

L114. The objective feels a bit woolly and non-descript as it is currently written. It would be good to explicitly mention each scheme to be used and reference.

L133 (and other places). Use of the term “components”. Component has a very specific meaning in oil and gas emission terms and refers to the smallest level such as valves. These are sub-site level measurements or process level measurements.

L140-142. It would be good to quantify what is sufficient wind, or what constitutes a negligible upwind source etc... in this list.

L157. This is key to the limitations of this study and is a real shame. I feel that the lack of operator buy in should be highlighted to point out how detrimental and hindering it is to scientific conclusions when the operator fails to provide activity data. I feel this should be highlighted to

make the point to policy makers that there should be a mechanism where by this can be requested within reason.

Table 2. Would it be possible to add details to each facility such as total emissions as predicted by inventory / nameplate capacity / age / gas throughput / any other production details?

Paragraph starting L201. As the choice here has some potential for human intervention / error – how much material difference does the choice being made?

L221-222. Description of SA algorithm application insufficient. Can this be referenced or if commercially sensitive please state.

L225. Remove last sentence as not adding anything. Would then move “Figure 3 provides...” to join to the previous paragraph.

L277. As the operators were not informed of this measurement it feels that this measurement is made under a very different sampling protocol, I feel this should be alluded to in the discussion, along with a note on how much pre-warning of measurements they were given for the flight measurements.

Figure 6. With Table 3, is this figure needed in the main text? Could it be in the SI? For table 3, would it be possible to add some information on inventory estimates so that a sense of “potential expected scale of emission” can be seen

L329. Can the reason for no further close flights be expanded on? Is it a safety issue or an operator choice issue?

L376. Can it be clarified that this is different from the emission conditions seen during the flights (i.e. single plume vs dual plume) or whether it was not possible to confirm? If that is the case this is good indication that emission conditions changed between sampling methods and demonstrates nicely the problems of scaling up spot sampling.

L431 (and please check throughout). Use of the term “error” when “uncertainty” is more correct. We do not know the true value so there cannot be known error as such.

L455. The error (see earlier comment!) estimate of 0.01% seems incredibly low. Can this please be checked

L504. Is there anything that can be said about what should be done if fundamental assumptions are broken? Is data collected under non-ideal conditions have any value for determining a flux measurement?