Response to comments from Referee #2, Kukka-Maaria Kohonen

We thank the referee for this thorough review. Careful consideration of the extensive comments and implementation of many of the suggestions have made this a stronger manuscript. Below, we address each question in turn. Questions and suggestions are in black, and our responses are in blue.

General comments:
Measurement methods in general require more description:

-Eddy covariance flux calculation description is lacking relevant information. The authors list as correction methods axis rotation, time lag compensation, WPL correction, and storage term correction. Which axis rotation method was used?

Response: More detail has been added to the manuscript (see lines 123-125 in Section 3.1). The standard double rotation (zeroing the average cross and vertical wind components) was applied (cf. Wilczak et al., 2001), and a planar fit method was tested, resulting in insignificant differences. As described in the response to Referee #1, in the test of planar fit method calculation, four sectors were defined: 286º - 76º (pond sector); 76º - 124º (east shoreline sector), 124º -259º (the south sector); 259º -286º (west shoreline sector). The resulting half-hour CH_4 EC flux and the original flux were within 0.0 ± 0.1 g m^{-2} d^{-1} of each other (mean and standard deviation of the difference). Therefore, as expected, during this campaign at this site the planar fitting method did not significantly change the final CH_4 EC flux results.

WPL correction should actually not be applied for this gas analyzer (Picarro G2311-f) as it is already included in the instrument itself.

Response: This was indeed an oversight on our part and we are grateful to the reviewer for pointing this out. We recalculated the fluxes in EddyPro without WPL correction, and found that the new results (half-hour series) are essentially indistinguishable from our original results, with an average decrease of 0.04% lower. Therefore, this correction issue had no significant effect on our results or conclusions.

Spectral corrections are not mentioned in the text. Spectral corrections (especially high frequency spectral correction) are essential in EC flux processing and can affect even the sign (direction) of the flux measurement. Recommended spectral correction methods are introduced in e.g. Aubinet et al., (2000) and Mammarella et al., (2009).

Response: Given the measurement height of 18m, spectral corrections are usually small, which is why we did originally not apply any spectral corrections. Recalculating the fluxes and applying a high frequency correction of low-pass filtering effects according to Moncrieff et al. (1997), we found the new numbers to be very close to the old results. On average, spectrally corrected values were 0.8% higher than uncorrected values. Therefore, this correction did not significantly affect the final pond emission results and conclusions.

Was u* filtering applied? If yes, what was the threshold and how was it determined?

Response: We carefully investigated this issue in our early analysis, and found that there was no evidence of fluxes becoming underestimated or erratic at lower u_* (Figure S4, original Figure S5). For most of the half-hour periods when the wind was from the pond, u_* > 0.1 m/s.

How about storage change fluxes, how were they calculated?

Response: Storage fluxes of CH_4 were calculated as the second term in equation (2), i.e.

$$F_{storage} = \int_0^z \frac{\partial c}{\partial t} \partial z$$

Eddypro assumed that the profile was linear from the measurement point to the ground and calculated the storage flux as a separate term. In this study, the storage flux was added to the calculated EC fluxes in the final EC fluxes. Given that the dynamic stability associated with pond wind directions was in the unstable regime 98% of the time, the storage correction made little difference to the net flux.

Fluxes from different wind directions are presented in this study, but it is not clear weather all these fluxes were processed in similar way. If all wind sectors are covered with different types of roughness elements (such as pond, buildings, trees), the different sectors should be processed (and fluxes calculated) individually. Environmental data required for the flux calculation (air pressure, temperature and humidity) are not described.

Response: Each half-hour flux was calculated independently, and cumulative/average statistics were calculated for the different wind sectors to ensure homogeneous conditions upwind for each sector. Meteorological inputs were described in lines 90-96.

-Gradient flux method has deficiencies.
Eddy diffusivity is calculated from CH4 EC flux, so gradient flux is not totally independent from EC CH4 flux measurements. I understand the eddy diffusivity is not taken directly based on EC measurements, but from a fit of Schmidt number against stability parameter. Even though making this fit makes gradient fluxes not directly dependent from EC, it should still be discussed how the usage of EC measurements in eddy diffusivity calculations affect the comparison between these methods, as it has not been currently discussed at all.
The authors refer to a study by Bolinius et al., 2016 where the eddy diffusivity is calculated from the heat flux measurements of the EC system instead of the gas flux. This is a well established method and I recommend the authors to study it more carefully and implement in their study as well. I suggest the authors at least compare this method to their original gradient flux calculations. Another study worth taking a look at is Rantala et al., (2014), where eddy diffusivity is calculated from the Monin-Obukhov similarity theory. Heat flux is independent from the gas flux, so calculating eddy diffusivity from the heat flux measurements will allow more reliable comparison between EC and gradient CH4 fluxes.

Response: We evaluated the method of using the heat fluxes to establish an eddy diffusivity KT early on, but found these diffusivities to be significantly noisier than those based on momentum. And an obvious problem with using heat fluxes as a baseline is that fluxes are very small at night, and therefore KT becomes very erratic and unusable. Km is also independent of the gas fluxes and has the advantages of being relatively well-behaved and continuous. There are similarities between our approach and that of Rantala et al (2014), but our approach does not rely on Monin-Obukhov Similarity Theory since we use the directly measured momentum flux, and the stability corrections that are explicit in the M-O approach are incorporated into our formulation of the Schmidt number.
We are well aware that our gradient fluxes are not truly independent of the eddy covariance fluxes. However, the fact that the Schmidt number we calculate agrees with previously published constant serves as an independent verification of the gradient flux approach. Even using a constant Schmidt number and stability corrections from literature, i.e. not using the measured EC methane fluxes at all, would have produced very comparable gradient flux numbers.

- Chamber measurements are currently not described at all but a proper method description is needed (what kind of chamber design was used, dimensions, how long enclosure time was after reaching equilibrium with carrier gas flow and inside air, how was the air flow implemented, how was the flux calculated, what kind of data selection methods were used etc.).

Response: The chamber measurements, which were performed by a third party independent of our project, followed the US EPA Standardized point measurement technique (adapted from Kienbusch, M., Measurement of Gaseous Emissions rates from Land Surface Using an Emission Isolation Flux Chamber, User Guide, EPA Users Guide, Contract No. 68-02-03389-WA18 (EPA/600/8-86/008), 1986. Regulations regarding chamber measurements in Alberta are given in https://open.alberta.ca/publications/9781460145814. The key steps are reproduced here:

1. Once the flux chamber (~0.1m^2 surface coverage area) is deployed on the target surface of interest, the valve of nitrogen cylinder was opened to begin purging the flux chamber with 99.9995 percent pure nitrogen gas. The flow rate of the nitrogen sweep gas was adjusted to a certain flow rate using the rotameter and this rate was maintained throughout the sample duration. The exhaust gas sample/purge rate did not exceed 2.5 L/min. This prevented ambient air entraining into the chamber and maintained a minimum exhaust rate 2.5 L/min out of the pressure equalization port. The GHG analyzer has an internal pump that operates at 0.5 L/min. The start time of the purge and the initial concentrations of CH4 were recorded on the field data sheet.
2. For the first 45 minutes, concentration readings were noted on the field data sheet in 15-minute intervals. Approximately five times the flux chamber volume was be purged in the first 45 minutes of sampling. Generally, the CO_2 and CH_4 concentrations reached steady-state 30 to 45 minutes into purging as indicated by a plateau in the real time data curve.
3. After 45 minutes, concentration readings were recorded every 10 minutes for a minimum of 75 minutes of total sampling duration. A minimum of 30 minutes of steady-state concentration data had to be obtained for the sample to be valid. The recorded concentrations and times on the field data sheet act as a back up to the GHG analyzer data files.

Comparison of fluxes is highly misleading and fundamentally flawed. The authors have included in the flux averages all data available, which are then compared with each other. What the authors should do instead is to select only those time periods/data points when all the measured fluxes are available, and then calculate averages that are comparable. If this is not done, it easily happens so that one of the methods is measuring e.g. more fluxes from one wind direction or time of the day than the other, which is causing a clear bias in the comparison.

Response: We have added text to clarify that the comparisons we show for the EC fluxes, gradient fluxes and inverse dispersion model fluxes are of course based on the set of simultaneous half-hour measurements over the 5-week study, when data was available to calculate all three. This does not apply to the comparison with the chamber fluxes, since the flux chamber measurements, conducted by a third party, happened to be performed when the wind was from the south and the micrometeorological methods (located on the south shore) could not observe the pond. For this comparison, the assumption is made that emissions from the pond are relatively time invariant during the period that was missed by the micrometeorological fluxes, as supported by the time series of fluxes for wind directions from the pond during the study period. This is in fact a common assumption made in many applications of flux chamber work, due to the snapshot nature of such measurements, and represents a significant limitation of flux chambers that we highlighted in lines 44-48 and in section 4.6 and 4.7. These well-known limitations were one of the reasons for exploring alternative methods for quantifying fluxes from such sources of fugitive emissions.

Conclusion section is currently an additional discussion section that should have been implemented in the section “Results and discussion” already. Proper conclusions – with no new information given but rather a summary with a perspective to future studies – is totally missing and should be included.

Response: A conclusion section has been added, and a comparison to previous results was inserted into Section 4.7.

Specific comments:

Table 1: This comparison does not make sense if the fluxes are not averaged from simultaneous measurements. You should only include the datapoints in averaging when you have a datapoint from all the methods. What does it mean that fluxes are "relatively steady"? The uncertainty estimation in footnote c is unclear.

Response: As explained above, the comparison of EC flux, gradient flux, and IDM flux is of course based on a set of simultaneous half-hour data points, when data was available to calculate all three, with wind directions from the pond. The exception are the flux chamber results, for which an assumption of time invariant fluxes during the concurrent micrometeorological flux time series is required.

The uncertainty estimation is based on a conservative, integrated approach encompassing all errors. In the real data time series, periods were identified when the flux did not fluctuate much, i.e. represented steady state conditions. In this case, we found in EC flux time series five periods of at least five half-hours with standard deviation not greater than 0.89. Then, the average of the standard deviations from these five periods was used as the uncertainty of EC flux in this study. For gradient flux and IDM flux, we used the same five periods, and calculated the average of the standard deviations from the five periods. This approach provides an upper limit or conservative estimate of the overall uncertainty in the final flux.

Table 2: Not clear what are the time periods for these flux estimates, should they even be comparable? Annual averages are different from summertime measurements. It would be interesting to see a comparison to natural waters or reservoirs as well, to see the high magnitude of the methane emissions.

Response: We are comparing all data available publically for this particular pond. The Small et al. (2015) data were from measurements in 2010 or 2011. Stantec report (2016) data were from measurements in 2013 and 2014. Baray et al. (2018) results were from the aircraft campaign in 2013, which was discussed in Section 4.7. These data were compared to provide context of results from this study. The comparison of annual averages vs. summertime measurements was discussed in Section 4.7, in the first paragraph. Natural lakes and indeed even wetlands emit at rates well below what we observed on this industrial pond, typically on the order of 0.005-0.05 g m^{-2} d^{-1} (Sanches et al., 2019).

Figure 1: It would be very helpful for the reader to include in the map the EC footprint lines and/or lines for approved wind directions. It is not very clear from the closeup image where exactly are the pond edges. Maybe this could be highlighted somehow? Add chamber measurement locations to this map.

Response: Yes, the EC footprint lines and chamber measurements have been labeled in the revised Figure 1. A sentence has been added to the caption explaining the different shadings of surface cover.

Figure 2: What is the correlation coefficient of the linear fit? How does it change if you use the original datapoints instead of binned averages for fitting? It does seem that data are very scattered with higher Km and Kc, how does this affect the fitting? What do the boxplots represent (what are the box limits, whiskers, center line etc)?

Response: The correlation coefficient for the linear fit was r^2=0.93. The original data points were indeed quite scattered, but produce the same slope. The figure caption has been revised to label box, lines and whiskers.

Figure 3: “Best fit” - determined by what criterion? The bins are not of equal size and I believe this is also affecting the fit. What do the boxplots represent (what are the box limits, whiskers, center line etc)?

Response: Bins are actually of equal size with bin width = 1, except for the last bin on the right for z/L > -0.34. At z/L=0.34, the exponential is equal to the Sc=0.923 found from the linear fit in Fig. 2; therefore, this was chosen as the point to switch from the exponential to the constant part of the function.

Figure 4: Fig 4b is not discussed anywhere and is a bit pointless without water temperature. In 4a, u* is missing interquartile ranges and 10% and 90% percentiles. In 4e sensible heat flux is missing quartiles off-pond and 10% and 90% percentiles on pond. Mark in the diurnal plots the times of sunset and sunrise to help the reader.

Response: All the information required is shown; from the temperatures in panel (b) and the temperature difference shown in (c), the absolute pond surface temperature can be inferred if desired. Panel (b) was in Figure 4 to show the typical diurnal cycle of ambient temperature. We have now cited panel(b) in Section 4.1.
The figure has been revised to include sunrise, sunset and off-pond heat flux quartiles.

Figure 5: Scale seems quite arbitrary, how was it defined? Directions are missing, where is north?

Response: The scale was set to include roughly equal numbers of data points in each range. The figure caption has been revised to label north.

Figure 6: You should add a, b and c to subplots. Colors of EC and gradient fluxes are too similar in the printed version and in the lowest panel red and green are used which is not color-blind friendly. You can check colorblind and printer friendly color choises e.g. from here: https://colorbrewer2.org

Response: Fixed.

Figure 7: Shade the pond area also here, similar to Fig. S6. What do the boxplots represent (what are the box limits, whiskers, center line etc)?

Response: Shade has been included here. Lower and upper bounds of the box plot are 25^{th} and 75^{th} percentile; the line in the box marks the median and the black square labels the mean; the whiskers label the 10^{th} and 90^{th} percentile.

Figure 8: What is the offset of the fit? It does not seem to be crossing y=0 at x=0 in neither of the plots.

Response: They both indeed cross (0,0).

Figure S2: What do the confidence intervals represent?

Response: The blue shade is 25^{th} percentile to 75^{th} percentile of the wind direction in degrees. This figure shows that the diurnal variation of wind direction is weak. However, this point can be made in the text without a supporting figure, as suggested by the other reviewer, and we decided to remove this plot.

Figure S3: “...countours of the EC footprint area”. It would be very helpful for the reader to get S3 b on top of a map, to see where the contours are crossing pond edges.

Response: We agree. The footprint contour is now superimposed onto the pond map in the revised Figure1. FigureS3 has been removed.

Figure S4 (now S2): It is not mentioned here which EC flux this is. Methane? Mention in each subplot which wind direction it is representing (in legend/title/xlabel/ylabel) to help the reader. What do the boxplots represent (what are the box limits, whiskers, center line etc)? Mention in the caption what is in each wind sector (pond, buildings, trees, etc).

Response: Yes, we mean methane EC fluxes. All these suggestions are accepted and Figure S2 has been updated. Lower and upper bounds of the box plot are 25^{th} and 75^{th} percentile; the line in the box marks the median and the black square labels the mean; the whiskers label the 10^{th} and 90^{th} percentile.

Figure S5 (now S3): Mention in the ylabel that this is methane flux. Mention in the caption what is the r2 representing (least squares linear fit?).

Response: Caption has been revised to note r^2.

Figure S6 (now S4): What do the boxplots represent (what are the box limits, whiskers, center line etc)?

Response: Caption has been revised.

Figure S7 (now S5): What do the boxplots represent (what are the box limits, whiskers, center line etc)?

Response: Caption has been revised.

Table S1: Are the fluxes compared here from exact same time periods? Same comment as for Table 1 about the uncertainty estimate and "relatively steady".

Response: Yes, the CH_4 gradient flux with variable Sc and constant Sc use exactly the same vertical mole fraction gradient data over exactly the same set of simultaneous data. An explanation of the uncertainty estimate was given in our response to comments on Table 1 above.

L10: "develop" is a little bit misleading here since the authors don't really develop any new method, rather compare already existing ones.

Response: While all three micrometeorological methods are of course well established, we are not aware of any previous instances of our approach of calculating the gradient fluxes through the use of a momentum flux diffusivity adjusted with a stability-dependent Schmidt number. Therefore we would like to keep the current term.

L11-12: Mention briefly which are these three flux methods in one sentence.

Response: Done.

L15: inverse dispersion model comes here from out of the blue. Describe it briefly before writing about the results.

Response: Done. The Inverse dispersion model is now introduced in line 12. A detailed explanation of the method is given in section 3.3.

L18-19: This sentence is a bit misleading. In one perspective it is quite obvious that a larger footprint represents a larger area. On the other hand if the EC tower is placed so that it is measuring only e.g. shallow area while actually the pond is deeper from a much larger area, then would EC be representative of the whole pond emissions? Then on the other hand nobody can know what is the real flux. It might as well be closer to the chamber flux than EC.

Response: There are several reasons that point to the eddy covariance fluxes being the more accurate estimate of the true fluxes. We have shown that fluxes were consistent for various wind directions across the pond, over a month of measurements, and they represent a large fraction of the pond surface. This is in contrast to the chamber measurements which cover a total of a few m^2 for instantaneous snapshot measurements, limited to regions of the pond accessible by boat. Implications have been discussed in the manuscript section 4.6 and section 5. Also of concern is the large interannual variability in flux chamber results, with 5.3 g m^{-2} d^{-1} in 2016, 2.8 g m^{-2} d^{-1} in 2017 and 11.1 g m^{-2} d^{-1} in 2018, despite similar operational conditions.

L21: Abbreviation AOSR is not used anywhere in the text

Response: Deleted.

L23: “Oil Sands” or “oil sands”? Throughout the manuscript.

Response: Made “oil sands” consistently throughout.

L48: “eddy covariance (EC)” and then use EC after this throughout the manuscript instead of eddy covariance

Response: Fixed.

L49: “area sources” or “source areas”?

Response: Area sources

L53: So only emissions can be measured with this method, not uptake?

Response: Uptake can of course be measured too, and would manifest itself as a negative flux. The two cited studies were emission flux studies.

L56: What is meant by “relatively well-defined spatially”? If the fluxes are well-defined, why do you measure them?

Response: We meant that the source area was relatively well-defined spatially, not that the fluxes were known. The ponds cover a well-known spatial domain, in a remote region far from urban activities and other sources.

L59: “Field study” is not a very descriptive title. Maybe “Site and measurement description”?

Response: Agreed & implemented.

L62: Trees are not part of natural landscapes? What is? How far were the other facilities? In the catchment area or further away? How large is the cathment area?

Response: This artificial pond is elevated above the surrounding landscape and has minimal catchment (~ 100m around the shoreline). The influx of industrial processed water vastly dominates the water budget of the pond. We added distances to the main facilities nearby to the text in lines 62-65.

L65: What is meant by "mobile tower"? How high were the measurements above water (which is more relevant than ground in the case of pond fluxes)?

Response: The tower mounted on a truck bed and can be easily towed from place to place for temporary installations. The base of the tower was less than 30cm above the water surface.

L69: Is this the diameter or radius? Inner or outer diameter?

Response: Outer diameter. This is now noted in line 73.

L71-72: I am not sure it can be said that turbulent flow is ensured. Reynolds number is _1300 according to my calculations, so it is possible that the flow is turbulent, but I wouldn't call it "ensured".

Response: Teflon tubing is generally labelled by outside diameter. Calculating the Reynolds number with the inside diameter of (3/8" minus wall thickness 1/16") D = 0.635 cm, with a flowrate of Q =117 $cm^3 s^{-1}$ and a kinematic viscosity of $\nu = 0.148$ $cm^2 s^{-1}$ gives Re = (Q D)/(ν A) = 4500, well within the turbulent regime.

L72-74: All kinds of measurements are presented that are not used in the analysis or shown anywhere in the manuscript. I suggest to leave out the description of those gas measurements not used in this particular study. Why is a 40 m long tubing required for 18 m height measurements? This will cause quite long lag time for EC. What are the three and four levels mentioned here?

Response: 40m tubing was used for all gradient levels including the 18m height measurement, to avoid systematic differences due to tube lengths. Three levels was a typo; the G2204 sampled from four levels, 8m, 18m, 32m on the tower plus 4m on the roof of the trailer. The length of the EC 3/8" OD line was 30m; this has now been added to the manuscript. We have removed the description of the G2401-M analyzer since its data was not used in this work.

L74: There must be some flush time of the tubings and analyzer between the different height measurements. How long is the flush time? One level cannot be measured 2.5 min during 10 min period if you take into account the flush time.

Response: Air was drawn through all 4 tubes continuously, and the only part of the flow system requiring flushing was the last 4m of tubing. To allow for the flow and pressure to equilibrate after each level switch, the first 30 s of the 150 s period at a given level were eliminated from the averaging process.

L76-77: Was there any drift of the instruments between calibrations? Did they compare well with each other?

Response: The calibration coefficient (slope) for CH_4 changed by 0.12% from before to after the study, and the offset by less than 0.002 ppm.

L80-82: This is well known EC theory and does not need to be explained.

Response: Since this information is fundamental to this paper and not all readers of this journal may be familiar with it, we chose to retain this.

L88-89: Was the infrared sensor calibrated somehow?

Response: No.

L92: How were the suitable wind directions determined?

Response: They were based on the map.

L104: EC also has its limitations, “benchmark” seems a bit exaggerated

Response: We have changed “benchmark” to “reference”.

L106: Response time and sampling frequency are not the same. Response time should be given in seconds, sampling frequency in hertz. EC measurements require both fast response times and high sampling frequency.

Response: Fixed.

L107: “CO2 and CH4 fluxes”

Response: Fixed.

L108-109: Reformulate the sentence. EC does not calculate anything, and in this case you are talking about gas fluxes explicitly (not e.g. heat flux since you mention mole fraction)

Response: “method” is inserted after “eddy covariance”.

L113-114: Repetition from above

Response: Modified.

L115: “storage change flux”. Out of curiosity, how large was the storage change flux? Often in lake studies they have been neglected but might be important as well.

Response: CH_4 storage fluxes was small. When the wind was from the pond, the storage flux was -2% to 3% (interquartile range) of the first term in equation (2).

L118-122: More description is needed on the processing methods used. How long was the lag time on average?

Response: More description has been inserted.
Covariance maximization method was used in time lag compensation. This method maximizes the covariance to variables (Fan et al. 1990), within a window of plausible time lags automatically calculated by EddyPro. The lag time on average was 12 second. (Fan et al., 1990).

L122: What do the different flags mean (what are the criterion)?

Response: As described in Mauder et al. (2016) and Mauder and Foken (2004), the quality test calculates the ratio of the standard deviation of CH4 flux to CH4 flux. Then, this ratio (or relative standard deviation) is compared to modelled results (as described in Mauder and Foken (2004)) to get a relative difference. The flags are determined based on this relative difference. In this study, we used the widely used overall flag system also described in Mauder and Foken (2004): flag = 0 when this relative difference < 30%, flag = 1 when 30% < this relative difference < 100%; and flag = 2, when this relative difference > 100%.

L123: "Gradient flux method"

Response: Fixed.

L130: units?

Response: Unit of Kc are m^2s^{-1}

L135: How do you define the gradient method footprint?

Response: This was mentioned lines 136-137 and explained in detail in the footprint section (4.2).

L168: Shifting winds are also a problem for EC measurements!

Response: That is correct. Since shifting winds are a problem for both methods, our way of excluding fluxes when the signs of EC flux and gradient fluxes were opposite at least partially excluded such situations from the comparison.

L177-180: What are the units of these variables?

Response: In this study for CH_4, unit of C and C_b is ppm, unit of Q is g $m^{-2}day^{-1}$. However, the formulation is valid for any consistent system of units desired.

L181: Why are L and u* used as inputs, since u* is already used in calculating L?

Response: To quantify stability, u_* by itself is insufficient, and the heat flux is also required, as incorporated in the Obukhov length L. Since L was already defined in the previous section, stating the inputs in this manner seemed the least confusing option to us.

L185: How close is "right beside"?

Response: Fixed. The new sentence " CH_4 mole fraction input was taken from the OP-FTIR measurement which was located 10m to the east of the flux tower."

L188-195: More description needed

Response: Yes, we have inserted more details. The full details are available through the Alberta online public report cited, and are not the focus of this paper.

L196-199: Not understood where this is used (which methods) and why. More description please.

Response: The standard method for flux chamber operations for compliance monitoring in Alberta was used. The details can be found in the online report cited in Section 3.4. Please also see the response above for the chamber measurements.

L202: “wind coming from” or “wind that came from”

Response: Changed to “coming from”

L214: Sunsets and sunrises are not directly seen from Fig 4d.

Response: Fixed. The revised Figure 4 now has yellow shades marking the range of sunrise and sunset times for the study.

L215-216: The same wind direction/ source area applies to gas fluxes. Why different wind direction analysis is applied to sensible heat fluxes and not the others?

Response: This section was talking about meteorological parameters. All fluxes, including sensible heat fluxes and gas fluxes, were analyzed in the same manner. Results in Table1 are based only on data associated with wind directions from the pond.

L229: You can only see this from Fig 4c. What is meant by “species”? Methane?

Response: This applies to all species emitted from the pond, not just CH_4.

L225-230: This is done for wind direction filtered data I assume? Should be clarified since all wind directions are analyzed in some way or another in this manuscript.

Response: This has been done for the all half-hour periods. The footprint polar plot shows the footprint under unstable conditions. This is noted more clearly now in the revised Figure 1 caption.

L230: This is not seen from Fig S3 (b), since you cannot see pond edges in the figure.

Response: Fixed, in the revised Figure 1.

L231: “gradient flux”

Response: Fixed.

L234: delete “however”

Response: Fixed.

L235-240: But there is much more data from off-pond direction than from pond direction. How does influence the analysis?

Response: That was the reality dictated by uncontrollable constraints such as site accessibility and weather, and one reason for scheduling a 5-week long campaign, to ensure a statistically significant number of days for most wind directions. There were 280 half-hour periods when the wind was from the pond and 98.6% of them were under unstable conditions, i.e. our footprint results have reasonable statistics.

L245-253: It is not clear why fluxes off-pond are reported since this is a study concentrated on pond emissions. Are these sectors processed in flux calculation individually or not?

Response: All the sectors were processed in the same way in EddyPro calculation. Off-pond fluxes were reported here to provide a measure both of the methodological noise in the signal, as well as the background (non-pond) flux magnitudes. Off-pond fluxes are either close to zero, or had a slight increase during the middle of day. The pond fluxes had no significant diurnal pattern.

L253-258: Are these now results from pond direction? Wind and turbulence are still driving the turbulent/diffusive transport of gases from pond to the atmosphere (e.g. Tedford et al., 2014).

Response: Yes, fluxes from pond direction. The pond flux diurnal cycle was shown in Figure S4(a). In the main manuscript we state "The lack of a diurnal variation of CH_4 EC flux observed when the wind was from the pond in this study was similar to the lack of a diurnal variation of CH_4 EC flux at another tailings pond reported by Zhang et al. (2019)."

L263-265: How do medians correlate? Take into account my earlier comments about representativeness as well.

Response: According to the numbers in Table 1, the median of CH_4 gradient flux is 20% lower than median EC flux. Again, the EC fluxes and gradient fluxes cover exactly the same periods and were calculated using the exact same measurements.

L270-272: This is quite far taken conclusion. Based on the results here you can only say that EC fluxes were used to calculate Kc, which of course then correlates well with EC.

Response: The idea behind this statement is that the K_c calculated in this way lets us calculate gradient fluxes for any species emanating from the pond that obeys the same physics (turbulent transport) and chemistry (inertness) as CH_4. Most gradient methods (modified Bowen, aerodynamic etc.) depend on some input of an EC flux, for example through u_* or the sensible heat, so there will always be some measure of autocorrelation. In our case, the link to the EC fluxes of methane are strictly through the parameterization of the Schmidt number, so a perfect correlation is not a given.

L272-274: And what were the outcomes of these studies? How do they compare to this study?

Response: These sentences were to show to compare our findings to the surprisingly few previous studies which compared EC flux and gradient flux of CH_4: "Zhao et al. (2019) compared CH_4 fluxes from an MBR method as well as from an aerodynamic flux model to EC fluxes for two small fish ponds, and showed that the MBR fluxes were well correlated with EC fluxes, with a mean 27% greater than the EC mean flux." Such studies are rare, so we feel our contribution represents a useful addition to the body of studies investigating EC, eddy diffusivity, and gradient fluxes.

L278-281: Medians are actually not that different and means are within confidence intervals. Perhaps different time periods were used to calculate the averages of the two Sc's?

Response: The average of two Sc methods used exactly the same set of half-hour periods, which are the entire period of this campaign. With stability z/L corrected Sc, the mean gradient flux is closer to mean EC flux, compared to using the constant Sc.

L292: Above it is mentioned that the footprints are similar but here that they are different? How would the different footprint of the concentration measurement influence the flux?

Response: For infinite homogeneous upwind fetches, the methods have the same footprint, given correctly placed gradient levels. In real-life situations, there can be differences, since the upper gradient point has a larger concentration footprint than the lower one, and therefore may see sources farther upwind.

L297: "Results of IDM fluxes.."

Response: Fixed.

L297-298: Where is this shown?

Response: It is shown in another AMTD manuscript: https://amt.copernicus.org/preprints/amt-2020-257

L300: What are bubbling zones and where are they located? It comes as a surprise here that the chambers are not measured on the footprint of EC. If you measure bubbling zones with lots of ebullition, how is the chamber flux calculated? Don't the bubbles bring sudden bursts of methane, invalidating the normal flux calculation methods?

Response: The locations of the 15 flux chamber measurements are marked in the revised Figure 1. As can be seen, most of them fall within the EC footprint. It is possible that the sudden bursts of CH_4 could invalidate the flux chamber calculation and lead to an underestimation of flux, as discussed in Zhang et al. (2019). We wrote this in line 334-337. Integrated over the footprint of micrometeorological flux measurements, the intermittent nature of ebullition will have a minimal effect.

L300-305: How are large the medians compared to average fluxes?

Response: The median of the 15 measurements is 2.3 g $m^{-2}day^{-1}$, and the mean is 2.8 g $m^{-2}day^{-1}$. In addition, the 15 measurement fluxes were scattered indicating "the pond was highly heterogeneous in terms of CH_4 emissions".

L332: replace "a month" with "five weeks"

Response: Fixed.

L336: Lower than what?

Response: Lower than results from three micrometeorological methods. Fixed.

L344: These are not comparable if not taken from same time periods and same footprints

Response: As explained in our response in the General Comments section above, the comparison between the chamber and micrometeorological flux measurements requires the assumption that emission rates from the pond did not change much in the few days before, during and after the chamber sampling. This assumption is supported by our CH4 flux time series. We included this comparison to put results of this study into context of historical data and current operational monitoring and reporting methods, and to shine the light on future monitoring needs, such as seasonal variability of tailings pond emissions. We acknowledge in the text that "This reflects a general complication when comparing the five weeks emission results in this study to annual emissions reported in the past."

L365: Excatly, different time periods are compared with each other making the method comparison useless in this form.

Response: Please see the response to the previous comment.

L383-386: In the equation there should be FCO2 and FCH4 instead of CO2 and CH4? These are results rather than conclusions. Abbreviation CO2eq is not defined and there are too many significant numbers in the result.

Response: Fixed.

Reference list: Two references are not peer reviewed yet, and there are quite many non-peer reviewed reports included.

Response: It is an unfortunate reality that there is very little information in the peer-reviewed literature on industrial fugitive emissions to the atmosphere in the Alberta Oil Sands, which is one reason for the importance of this current manuscript. This lack of published information makes it difficult to avoid referring to grey literature. We have removed the two references to the yet unpublished manuscripts by Moussa et al., and updated the information on the You et al. manuscript on FTIR measurements in AMTD.

References

Fan, S. M., Wofsy, S. C., Bakwin, P. S., Jacob, D. J. and Fitzjarrald, D. R.: Atmosphere-biosphere exchange of CO_2 and O_3 in the Central Amazon Forest. Journal of Geophysical Research, 95: 16851-16864, 1990.

Mauder, M., and Foken, T.: Documentation and instruction manual of the eddy covariance software package TK2, https://core.ac.uk/download/pdf/33806389.pdf, 2004.

Mauder, M., Liebethan, C., Göckede, M., Leps, J-P, Beyrich, F., and Foken, T.: Processing and quality control of flux data during LITFASS-2003, Bound-Lay. Meteorol., 121, 67-88, 2006.

Moncrieff, J. B., Massheder, J. M., de Bruin, H., Ellbers, J., Friborg, T., Heusinkveld, B., Kabat, P., Scott, S., Soegaard, H., and Verhoef, A.: A system to measure surface fluxes of momentum, sensible heat, water vapor and carbon dioxide, Journal of Hydrology, 188-189, page 589-611, 1997.

Sanches, L. F., Guenet, B., Marinho, C. C., Barros, N., and de Assis Esteves, F.: Global regulation of methane emission from natural lakes. Scientific Reports, 9, (1), 255, 2019.