Response to comments on the revised manuscript, December 2020

The authors have made some improvements to the previous version of the manuscript and answered satisfactorily to some of my previous comments, but the bigger comments from last comment round were not taken into account properly. There are still four major issues (in addition to few smaller ones) that require addressing before publication in AMT.

Response: We are very grateful for the reviewer's persistence, which prompted us to re-examine our analysis and to discover two issues that we have now corrected. The first was incorrect accounting for the delay time of the gas concentration signal relative to the sonic anemometer, the second was incorrect application of spectral corrections. The combined corrections resulted in an increase of the fluxes by 25%. Details are given below. Since these adjustments in the eddy covariance fluxes were not constant in time and therefore do not translate into a simple scaling of the gradient fluxes, changes in most of the figures will be noticed, as well as in the comparisons between methods.
We also thank the reviewer for insisting on additional statistical tests to determine the significance of the comparisons. These tests made it clear that the gradient flux method is not very strong at predicting fluxes from one half-hour to the next, and that averages of fluxes binned with by wind direction sectors or by hour of day)is required to produce statistically significant agreement. We have moderated our statements in the abstract and conclusion accordingly.

1) The manuscript and its analysis would really benefit from statitistical tests used to check wether the fluxes measured by different techniques really differ or not. The amount of chamber measurements is probably not enough for this purpose, but fluxes from EC, IDM and gradient flux methods could certainly be used for such tests. You should also test if the nighttime and daytime CH_4 fluxes really are statistically the same or not.

Response: we thank the reviewer for this suggestion. Two-sample paired t-tests were applied to the half-hour fluxes or average fluxes binned with wind direction sectors from EC, IDM and gradient methods, when the wind was from the pond.

Table 1: t-test of fluxes from three methods when the wind was from the pond

Variable 1	**Variable 2**	**p-value (half-hourly)**	**p-value (binned by wind direction sectors)**
EC flux	Gradient flux	0.003	0.30
EC flux	IDM flux	$< 10^{-15}$	0.08
Gradient flux	IDM flux	$< 10^{-4}$	0.33

To test if there was a statistically significant difference between daytime and nighttime fluxes, half-hour fluxes were separated by day and night, defining day as 6:00 to 21:00, and night as 22:00 to 5:00 MDT (local time). With t-test results on average fluxes binned with wind direction sectors, day average fluxes and night average fluxes are statistically the same.

Table 2: t-test of fluxes from three methods when the wind was from the pond, fluxes were binned by wind direction sectors and weighted by pond sector area.

	Flux during the day (mean)	**Flux during the night (mean)**	**p-value (half-hourly)**	**p-value (binned by wind direction)**

EC flux	7.4	8.2	0.16	0.54
Gradient flux	7.2	7.6	0.04	0.94
IDM flux	4.9	7.6	$< 10^{-5}$	0.04

We have modified the text to reflect that (1) EC, gradient, and IDM mean fluxes are statistically different on the half-hour scale, and statistically agree better after averaging with wind direction sectors, when the wind was from the pond; (2) fluxes during the day and at night are statistically not different, although there was difference in IDM fluxes between day and night time

In Section 4.3, we modified the first sentence in the second paragraph: There was no statistically significant diurnal pattern of the CH_4 EC flux when the wind came from the pond direction (WD ≥ 286°, or WD ≤ 76°) (relative standard deviation is 15%, p=0.54) (Fig. S4 (a)).

In Section 4.4, we modified the text to read: "Due to significant scatter, the half-hour gradient fluxes were statistically different from the EC fluxes when the wind was from the pond direction (p-value=0.003). They were moderately correlated (slope=0.80, r=0.32, Fig. S7(a)). To obtain some comparability, it is therefore necessary to average blocks of data into appropriate bins. A t-test of the gradient and eddy average fluxes binned by wind direction (22.5° blocks) yielded a p-value of 0.30, and hourly diurnal averaged fluxes agreed with a p-value of 0.09. The pond area weighted mean gradient flux was 8% less than EC flux, and the median was 18% less than EC flux (Table 1)."

In Section 4.5, we modified the text to "IDM and EC flux showed reasonable correlation (r=0.62) with a slope of 0.69 (Fig. S7(b)), although the averaged half-hour IDM fluxes are significantly different from EC fluxes ($p<10^{-4}$). Binning into 16 wind direction sectors similar to described in Section 4.4 yielded agreement at the p=0.08 level. The pond area-weighted mean IDM flux was 30% smaller than EC flux, and the pond area-weighted median IDM flux was also 30% smaller than the EC median flux. The IDM flux showed weak diurnal variations when the wind came from the pond directions (Fig. S8), with smaller fluxes during the day, compared to fluxes at night (p = 0.04), inconsistent with EC and gradient fluxes. As stated in Section 3.3, half-hour periods when u_*<0.15 m/s were excluded in IDM calculation (Flesch at al. 2004). This filtering excluded more nighttime fluxes than daytime fluxes, which caused more limited data in IDM nighttime fluxes and biased the t-test."

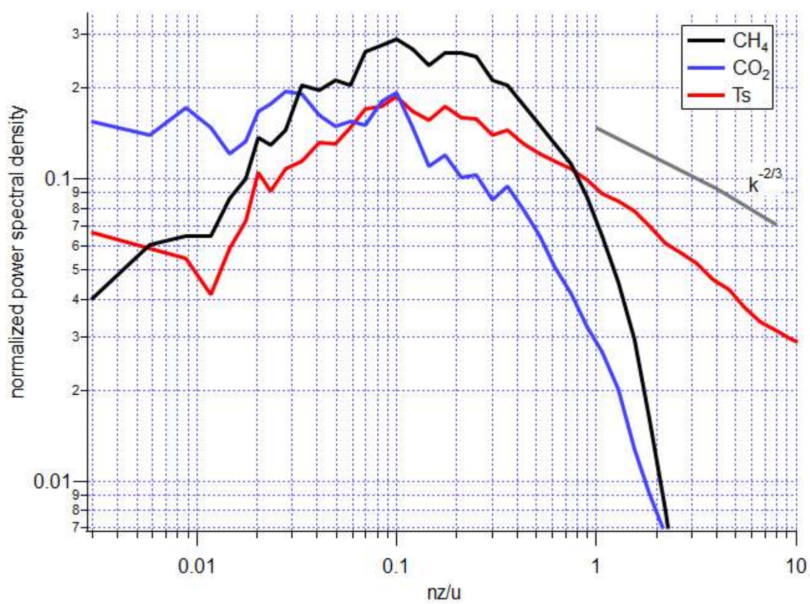
In the Conclusion, we modified the text: "The gradient and inverse dispersion methods agreed moderately with EC results (18% and 30% lower, respectively)".

2) About spectral corrections of the flux measurements. The authors responded to the request of spectral corrections that the high frequency spectral correction is not important at a measurement height of 18m and changed the average flux by 0.8%. While it is true that the importance of smaller eddies decreases at higher measurement heights, they still cannot be neglected at 18m (not high enough and smaller eddies probably still exist). In addition, the importance of low frequency spectral correction increases with higher measurements, and should be accounted for in the analysis. Why are the spectral corrections still not included in the measurement description? What kind of spectral correction method(s) did you use? You should check the power spectra of your measurements to check if high- and/or low frequency spectral corrections are needed. I am not conviced by only comparing the average flux, without low

frequency correction. If you still are omitting the spectral corrections, you should justify it in the text very clearly and include the power spectrum and cospectrum that proves it.

Response: We are thankful for the reviewer's persistence, since a more careful review of the application of previous spectral corrections uncovered some problems with our usage of EddyPro, which we have now addressed. Cospectral analysis indicated that rather than the minimal effect of < 1% we previously stated, spectral losses actually accounted for an average of 10%. A plot has been added to the supplemental information (S1) to show the average normalized cospectral densities for CH_4, CO_2 and sensible heat for those periods when the wind was from the pond and the data quality flag for the CH_4 flux, according to Mauder and Foken (2004), was either 0 (best quality) or 1 (good quality). Rather than plotting the cospectrum as a function of a normalized frequency (nz/u), we chose to use the natural frequency since high frequency losses will more likely be directly tied to natural frequency. Also, we chose a linear y-axis to facilitate the estimation of losses, which in this format are proportional to the missing area under the curve. High frequency corrections were applied using the approach outlined in Horst (1997). Low frequency spectral corrections according to Moncrieff et al. (2004) were applied, but did not significantly affect the CH_4 flux, as can be deduced from the relatively clean cospectral shape and similarity to the sensible heat cospectrum below 0.01 Hz (Fig. S1). It is also clear that the same cannot be said for CO_2, which is more frequently affected by signals varying over time scales greater than ~ 10 minutes and less likely to be due to turbulent flux from the (pond) footprint. This is also a good confirmation that the pond is the dominant source of essentially the complete CH_4 cospectrum, and that background fluctuations or sources outside the footprint are unlikely to significantly affect the CH_4 fluxes.

The power spectra, shown here, also indicate that a high frequency drop-off for the gas measurements in the inertial subrange faster than that of temperature, starting at normalized frequencies as low as 0.3, corresponding on average to a frequency of 0.06 Hz:

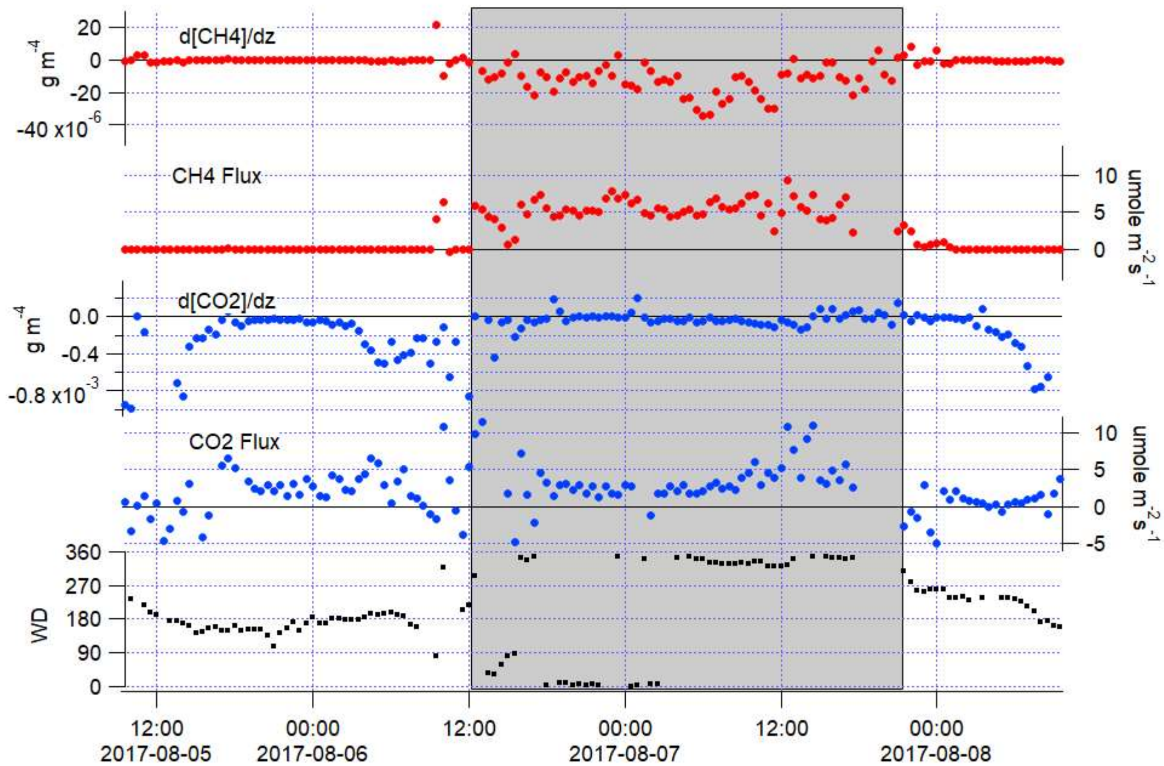


To provide a summary of the spectral analysis and corrections performed, we added the following text to section 3.1, last paragraph:

"Covariance spectra were examined for signal losses at higher frequencies (smaller eddies) during transit of the sampled air through the sample line, finite sample cell volume, and instrument response (Fig. S1), accounting for a loss of typically 10% of covariance signal, compared to the sensible heat cospectrum that does not suffer from equivalent losses. Spectral corrections following Horst (1997) were applied to correct for these losses. Low frequency losses at the low frequency end of the spectral peak due to the finite averaging time were applied according to Moncrieff et al. (2004). The EC flux quality flag was categorized into 3 classes: 0 (best quality), 1 (good quality), and 2 (poor quality) (Mauder et al., 2006; Mauder and Foken, 2004). Only EC fluxes with flag 0 or 1 were included in further analysis."

3) The fact that the gradient flux measurements are relying heavily on EC CH4 flux measurements is still not discussed with the flux comparison results. I well understand that the gradient flux was calculated based on a fit to CH4 EC fluxes and not directly to the fluxes, but there is still a strong link. What would make it a bit more reliable is to test the methods I recommended in the previous comments or make the same fit you have now done but using CO2 fluxes instead of CH4, and then calculate the gradient fluxes in similar fashion. In any case, it should be discussed in the results how your calculation method affects the comparison!

Response: The CO_2 flux, as seen in the average cospectrum in Fig. S1, was not nearly as well-defined at the CH_4 flux in this location, due to significant diurnal background variability as well as the presence of various CO_2 sources and sinks in the surrounding area, whereas for CH_4 there was only one dominant source (as supported by the wind direction dependency as well as the cospectra). This is what made this location a good place to relate gradients to gradient fluxes, using CH_4. Also, the CO_2 fluxes out of the pond were, relatively speaking, significantly weaker than the CH_4 fluxes, as can be illustrated with this figure showing about 3 days of data:



The CO_2 gradients for non-pond wind directions are significantly larger than those for pond directions, and the signal-to-noise ratio of the CO_2 flux significantly lower than for CH_4. Regardless, we have now added the following paragraph to the paper in section 3.2:
"It is possible to calculate K_c values based on CO_2, in order to avoid potential circularity arguments when calculating gradient fluxes of CH_4 using this approach. However, the CO_2 flux signal from this pond was confounded by the strong natural variability of the CO_2 background, and the smaller signal-to-noise ratio of the pond CO_2 flux compared to the CH_4 flux (Fig. S1). Regardless, K_c values based on CO_2 were calculated, and found to be noisier but statistically not different from those based on CH_4 (t-test p-value= 0.09, based on fluxes binned into 16 wind direction sectors). It would also be possible to base the calculated K_c values on the sensible heat flux instead of the momentum flux, but due to the absence of significant heat fluxes at night, this would not provide the continuity that the momentum fluxes afford."

4) Last but not least, many of my previous comments were either ignored (e.g. text is not well organized: in methods section EC and gradient instrumentation descriptions are not well separated, in the results on-pond and off-pond results are mixed and the reader easily gets lost, spectral corrections are still ignored, gradient-EC (in)dependence still not discussed...), not applied everywere in the manuscript (for example eddy covariance → EC, including medians together with means, some clarifications, e.g. about tube dimension), or were answered in the author response, but not applied in the revised manuscript (e.g. discussion on shifting winds affecting gradient fluxes are also affecting EC fluxes, medians not included in the text even when they were in the response). Many of the comments listed below are the same as before, as they were not implemented in the revised manuscript. Overall, it seems not much attention was given to the revision of the manuscript (e.g. a figure is referred that does not exist).

Response: We appreciate the reviewer pointed out these points potentially causing confusion. We did not understand the point about including median together with means in the first review since there was some confusion about line numbers. We have now put the median flux of flux chamber results in our revision. All the points mentioned above were addressed individually in the following detailed comments and in the revised manuscript.

Detailed comments (line numbers refer to line numbers in the "track changes" version of the manuscript):

Table 2: What are BDL and NA (mention in caption)

Response: fixed.

Fig 2: Write the correlation coefficient in the plot.

Response: fixed.

Fig 3: What I meant by my previous comment, was that the bins are not of equal size in terms of number of data points included in each bin. And that is why the fit does not really follow the original datapoints. How does the fit change, if you use equal number of datapoints for each bin? What is the fit equation?

Response: There was an error in the figure caption of Figure 3. The yellow points are the Sc of the entire data, including pond directions and non-pond directions. The boxes and fittings were on Sc of pond direction only. To make this clearer, the figure caption is revised, and in the revised Figure 3 the Sc points when the wind was from the pond direction are highlighted in black. Please note, since EC flux changed after the spectral correction, the calculated Sc values for each half-hour also changed, so the data in the revised Figure 3 are different than the original points. To reflect the change, the green line showing the original fitting with the original data points are included in the figure below for reviewers to see.

We also accepted reviewer's suggestion to bin data into bins with an equal number of points, and performed the same fitting with the median of each 10-point bin. The medians of each bin are shown as blue circle (z/L is the median z/L of each bin), and the new fit is the blue solid curve. In addition, we also tried our original way of binning data, and fitting results are shown by a red solid line. Fitting results of these two binning approach are close. We have accepted reviewer's suggestion, and only results from the approach of binning every 10 points (blue line) is used in the final results in the revised manuscript. The equation of the new fit:

$$S_c = \begin{cases} 0.08 + 3.13 \times 10^{-9} e^{\left(\frac{\frac{z}{L}+19.5}{1.008}\right)}, & \frac{z}{L} < -0.18 \\ 0.74, & \frac{z}{L} \geq -0.18 \end{cases}$$

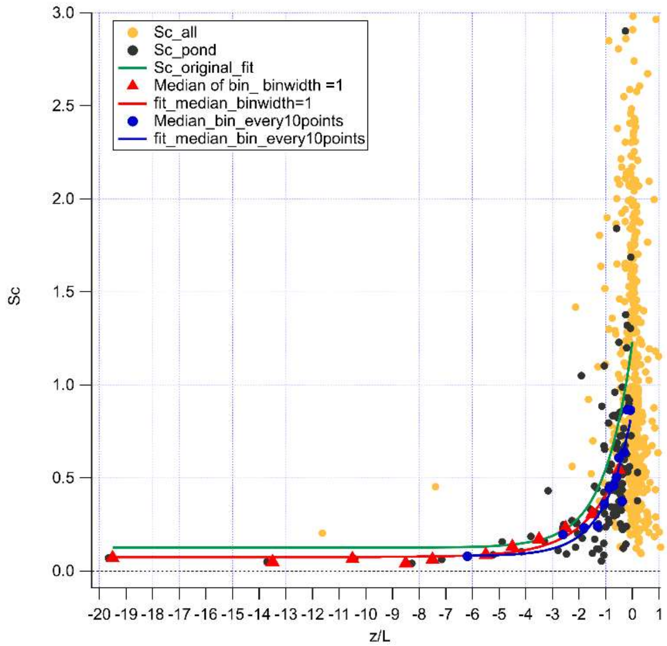


Fig 5: I still advice to use a uniform color scheme here, as the present one is highlighting differences in the range of 1.9-3 ppm, which are not really as dramatic as the ones from 1.9-9 ppm. My suggestion is to use a uniform color scheme, e.g. similar colors as you have used in scales 3-9 ppm. I recommend to read a recent Nature Communications article about colormap choises (https://doi.org/10.1038/s41467-020-19160-7). Adding the radius lines (0.2, 0.4, 0.6, 0.8, 1) on top of the wind rose would help the reader a lot (same goes for Fig S1).

Response: fixed. Please note, the original Figure S1 has become Figure S2.

Fig 6: The Xtick labels are too close to each other. Widen their distance or make the labels e.g. in 45 degree angle. You can also leave out the hours, just put dates in each xtick and specify in the caption that date tick represents midnight. Why is the lowest panel missing horizontal grids, when other panels have them?

Response: fixed.

Fig 7: The shaded areas are not the same that you mention in the text!

Response: thank you for catching this. The shades have been modified to only include wind from the pond sectors accounting for a safety margin, i.e. WD ≥ 286º or WD ≤ 76º. The original shades included sectors covering the shorelines.

Fig S5: Somehow mark in the figure daytime/nighttime.

Response: fixed. The sunrise and sunset timing are labeled.

Add the footprints of stable/unstable/neutral conditions to the supplement.

Response: This has been inserted into the supplement as Figure S3, and mentioned in the revised manuscript.

L18-20: Larger footprint together with frequent sampling.

Response: We modified this sentence as: “These results indicate that the larger footprint together with high temporal resolution of micrometeorological methods results in more robust emission estimates representing the whole pond.”

L29: oil sands

Response: fixed.

L56: Not only emissions, but also uptake, can be measured with the technique. Change to "gas fluxes" or "surface-atmosphere exchange"

Response: fixed.

L59-60: The sentence still reads that the fluxes are well defined spatially. Reformulate the sentence.

Response: We have reformulated this sentence:
“Tailings ponds represent a useful testing ground for a multi-method comparison of flux measurement techniques due to their reliability as sources of significant fluxes, relatively well-defined sources areas, and minimal other anthropogenic sources in the immediate vicinity.”

L74: 1/2" inner or outer diameter?

Response: fixed. “outer diameter” has been inserted in the text.

L77: H2O not defined. Mention that these are used for EC measurements

Response: fixed.

L78: Even with the numbers and formula you gave in the response, I get the Reynolds number ~1600. I don't understand how you get 4500.

Response: The Reynolds number is indeed only around 1600. In light of this, we have removed the sentence about plug flow. We apologize for this mistake, which we cannot reproduce now, and thank the reviewer again for her persistence. Upon revisiting the flow and resulting delay times, we discovered that our settings for the automated delay calculation (through cross-covariance peak detection) in EddyPro were incorrect and resulted in incorrect delay times, affecting 54% of all data points. Fixing this increased the covariances on average by about 15%. Adding this to the spectral corrections explained above yielded an overall average increase in the fluxes by about 28% for CH_4, and 35% for CO_2.

L73-78: Lots of confusion about sampling tube lengths. In the original manuscript gas concentrations were measured through 45 m tube and EC through 40 m tube. In the response you mention 40 m for gradient and 30 m for EC, in the revised manuscript gradients are measured with 45m and EC with 30 m. This should be quite straightforward and well documented...

Response: Sorry about the confusion. The eddy covariance tube was definitely 30m long, and all the gradient tubes were 40m in length.

L89: "Friction velocity (u *) can also be calculated from measured u, v, and w.". There is no point to this sentence if you don't present the equation, or tell how it is calculated.

Response: The equation of u_* has been inserted there in the revised manuscript.

L116: EC has already been defined....

Response: fixed.

L120: "which in this study limits the method to sensible and latent heat (H2O) fluxes, momentum, CO2 and CH4 fluxes." Since you don't show or describe other measurements, this sentence is unnecessary

Response: we indeed have sensible heat fluxes shown in Fig. 4 (e) and have used momentum fluxes in equation (4) and the CO2 fluxes in Table 2 and section 3.2 and 4.7, so we would like to retain this sentence.

L121: eddy covariace -> EC! Check this everywhere in the manuscript!

Response: All the “eddy covariance” have now been replaced with “EC”.

L121: replace "..the eddy covariance method simply calculates the flux by averaging.." by "in the EC method, flux is calculated by averaging…"

Response: fixed.

L126-130: Mention that you assume linear concentration profile for storage change flux calculation

Response: implemented.

L132: Mention the average time lag in the text.

Response: implemented.

L130-135: Mention that each wind sector was processed individually and how you took into account the different roughness elements of different wind sectors.

Response: we inserted this detail at the end of Section 3.1.

L190: Again, shifting wind directions are also a problem for EC, not just gradient fluxes! So these occasions are not probably due to shifting wind!

Response: Correct, they would also be a problem for EC. We have eliminated this sentence, since upon re-examination it contained circular reasoning and didn't actually describe what we did. A speculative reason for such reversed gradients would be a surface layer out of equilibrium with the pond, with elevated CH_4 concentrations being observed at 32m relative to those at 8m, since the concentration footprint at 32m is significantly larger than that at 8m. As we mention in the last paragraph of 4.2, the gradient footprint is equivalent to the eddy covariance footprint at the geometric mean height of the two gradient levels, but the underlying assumption is an infinitely homogeneous fetch. If the 32m concentration footprint reaches beyond the pond limit, this assumption may be broken.
In our revision, we slightly modified our process, and kept some of the negative gradient fluxes. Due to the scatter natural of the half-hour gradient fluxes, we excluded outliers of the lowest 2.5% and the highest 2.5% of data. In this way, we did not just eliminate negative gradient fluxes. We have stated this process clearly in the Section 3.2: "To lessen the impact of extreme outliers, the final pond average fluxes reported were based on gradient fluxes between the 2.5th and 97.5th percentiles".

L200-203: Mention the units you have used!

Response: implemented.

L218: I looked the report cited here and still did not find how the chamber fluxes are calculated. Please provide a formula for flux calculation. I understand this may not be the focus of this paper, but chamber fluxes are quite sensible to the calculation method used, and thus I think it is valuable information for the reader.

Response: Sorry, this time we understand what was being asked. The report of Alberta Environment and Parks Section 3.6 has included the EPA method. To implement this comment, we included the EPA (1986) report and cited the key equation (3-5) in the revised manuscript as equation (8).

https://nepis.epa.gov/Exe/ZyNET.exe/930013RX.TXT?ZyActionD=ZyDocument&Client=EPA&Index=1986+Thru+1990&Docs=&Query=&Time=&EndTime=&SearchMethod=1&TocRestrict=n&Toc=&TocEntry=&QField=&QFieldYear=&QFieldMonth=&QFieldDay=&IntQFieldOp=0&ExtQFieldOp=0&XmlQuery=&File=D%3A%5Czyfiles%5CIndex%20Data%5C86thru90%5CTxt%5C00000029%5C930013RX.txt&User=ANONYMOUS&Password=anonymous&SortMethod=h%7C-&MaximumDocuments=1&FuzzyDegree=0&ImageQuality=r75g8/r75g8/x150y150g16/i425&Display=hpfr&DefSeekPage=x&SearchBack=ZyActionL&Back=ZyActionS&BackDesc=Results%20page&MaximumPages=1&ZyEntry=1&SeekPage=x&ZyPURL

L219: N2O not defined

Response: fixed.

L220-223: Again, I still find this section quite confusing. Is this then the final flux you use in the flux comparison, or something else? Which flux measurements are used for this? EC, gradient, chamber, IDM? Provide more details.

Response: Yes, these are the final average fluxes representing the pond emission during this study for each of the methods (EC, gradient, and IDM). Section 3.5 has been modified: "The area weighted averages of fluxes results are summarized in Table 1 and serve as the final average fluxes representing the whole pond over the study period."

L303-308: Again, it is not clear wether you are now describing the off-pond or pond fluxes. For pond fluxes, wind should play at least some role, enhancing the turbulent transport of gases from the pond. Specify in the text if you are focusing on pond or off-pond fluxes in this discussion, since in the previous paragraphs you are describing both. As this indeed is from pond direction only (as you state in the response), discuss why they should not play any role, i.e., why more mixing would not bring up methane produced deeper in the pond. Looking at Fig. S3, even though there is no clear linear relationship, it is still clear that lowest fluxes are not measured at high wind speed and highest fluxes are not measured at lowest wind speeds. So there is some kind of relation to wind.

Response: The r of the linear regression of EC flux (from the pond direction) and wind speed at 8m is 0.4, so we modified the text: *"Relationships between the flux when the wind was from the pond and various meteorological parameters were investigated, and results show that fluxes showed weak dependence of wind speed, u_*, water surface temperature, or the temperature difference between the water surface and 8 m (Fig. S5)."*
It is of course physically possible that higher wind speeds could enhance mixing of the water near the water-air interface, as has been observed and parameterized elsewhere ***(cf. Cole, J. J. and Caraco, N. F.: Atmospheric exchange of carbon dioxide in a low-wind oligotrophic lake measured by the addition of SF_6, Limnol. Oceanogr., 43, 647–656, 1998)***. Visually we saw little wave formation even during windy periods, suggesting that the chemical composition, and possibly the presence of surface films, suppressed transfer of momentum from the air to the water. Also, it is unlikely that the production rate of methane by microbes in the lower (anoxic) strata of the pond are affected on a short (< daily) time scale; in other words, the source strength is likely mostly independent of wind speed, even if the transport mechanism varies. We modified the text as follows:
"Relationships between the flux, when the wind was from the pond, and various meteorological parameters were investigated, and results show that fluxes showed weak dependence of wind speed, u_*, water surface temperature, or the temperature difference between the water surface and 8 m (Fig. S4), i.e. they were not major drivers of the CH_4 emission rate. CH_4 at this site is mainly produced through the methanogenesis of hydrocarbon by the microbes in the fine tailings covering a range of depth in the pond (Penner and Foght, 2010; Siddique et al., 2011;Siddique et al., 2012), and therefore is not directly affected much by the meteorological conditions at the surface or above the pond."

L324: Report also the median flux differences, as requested before.

Response: implemented.

L327-338: You are still not discussing the relation to EC fluxes. Yes, the Kc was determined from a fit made to EC fluxes, and yes, almost all gradient flux methods require some input of the EC system. But how would the results look if you determine Kc from CO2 flux instead of CH4 flux? That would lead to at least a bit more independent comparison to EC CH4 flux. By minimum, you should discuss how the derivation of Kc from EC CH4 flux is affecting your comparison, or justify why it is not affecting at all.

Response: Please see our response to major point 3) above. Calculating K_c from the CO_2 data produced noisier results for the reasons discussed above, but the results were statistically not significantly different from those for CH_4. In other words, using the K_c derived from CO_2 would give us similar gradient fluxes, but they would be even noisier than those from CH_4. Dividing the eddy flux by the gradient (measured by a separate instrument, incidentally) results in a diffusivity that represents the transport of a nonreactive gas in general; we chose the best-resolved gas (in terms of signal-to-noise) to calculate it, and that happens to be CH_4 in this location. Our approach then hinges K_c on the stability (z/L), thereby removing it another step from direct correlation with the EC flux used to calculate the $K_c = f(Sc)$ relationship. The magnitude of the gradient flux is a direct function of the EC flux, thereby ensuring that the gradient flux will on average be about the same as the EC flux. An independent verification of the approach is given by our *Sc* values being similar to the few numbers in the published literature.

L335: Which studies? Add references to the sentence.

Response: we meant studies discussed in the sentences following that sentence. A study with another tailings pond (Zhang et al. 2019), and studies with other water surfaces (Schubert et al. 2012; Podgrajsek et al. 2014; Erkkilä et al. 2018). In the revised manuscript, we have inserted those references at the end of this sentence.

L340: Again, you cannot really say they are lower when the medians are almost the same, and also meand within the confidence intervals. It would be good to try some statistical tests to see if they really differ or not. Add it to the discussion.

Response: A t-test was performed with half-hour gradient fluxes with the both the variable and constant Sc approaches and the result was $p < 10^{-21}$, indicating the two results are indeed statistically different. The pond median flux with the constant Sc approach was 34% lower than the median of fluxes with the variable Sc approach, and the mean was 33% lower. We have implemented this comparison in the main text in the last paragraph in Section 4.4: "Gradient flux calculated from a constant S_c were significantly lower than gradient fluxes with the variable S_c approach ($p < 10^{-21}$, pond average mean/median is 33%/34% lower)."

L351: There is no Fig 9... You might mean Fig. 8b.
Response: Correct, thank you for catching this. Also, we have decided to move Fig. 8 into the supplement as Figure S7.

L366: Again about the bubbling zones.
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 CH4 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."
I did not find any discussion on lines 334-337 on the subject (on the previous manuscript, on the revised manuscript, or the author response).

Response: This might be the line number issue. We have had this discussion in all the versions since the original draft. In the middle of Section 4.6 when we first discussed Zhang et al. (2019), we have "Zhang et al. (2019) measured CH_4 emission from another tailings pond, and reported flux chamber measurements were more than 10 times greater than fluxes from the EC method. They stated that strong eruptions of bubbles could overwhelm the chamber to result in a local underestimation of the flux. On the other hand, the lower EC flux estimate suggests that the area average flux was being overestimated by extrapolation from the chambers, which may have preferentially been located over bubble zones. Their EC fluxes were two orders of magnitude smaller than CH_4 flux in this study. Results from this study and Zhang et al. (2019) suggest that average tailings pond CH_4 emission extrapolated from a few individual flux chamber measurements may significantly underestimate or overestimate fluxes relative to area-averaging micrometeorological measurements."

L405: EC

Response: fixed.

L416: As a side note, Pond 2/3 is a great name for a pond!

Response: There are historical reasons for the name. There used to be adjacent ponds numbered "2" and "3", but they were merged in the 1980s and to retain this history in the name, Suncor labelled the new unified pond "2/3".

L459: Abbreviation CO2eq is (still) not defined

Response: fixed. "(F_CO_{2eq})" is inserted into that sentence after "the equivalent CO_2 flux"

Somewhere in the discussion, mention how high these CH4 emissions are compared to natural (wetland/lake/pond) emissions from other studies, just to give some idea of the flux magnitude.

Response: we have inserted the following at the end of the first paragraph in Section 4.7.

"Natural lakes and wetlands emit at rates typically on the order of 0.005-0.05 g m^{-2} d^{-1} (Sanches et al., 2019)."

L462: Mention the methods used.

Response: implemented. We have modified this sentence:

"Results in this study have provided several estimates of the emission of CH_4 from this tailings pond using EC, gradient, and IDM methods,…"

L466: "in 2017" is not needed here

Response: deleted.

L467-468: "micrometerological flux measurements"

Response: fixed.

L468: "larger footprint together with high temporal resolution"

Response: fixed.

L469: To be accurate, the measurements are still not representing the whole pond. But they are representing most of the pond area. Reformulate the sentence.

Response: That is true. The modified sentence states: “The better agreement between the three micrometeorological measurements flux results suggests that the larger footprint of micrometeorological measurements results in more robust emission estimates representing most of the pond area.”

L471: Further studies of what? EC, chamber, IDM, gradient, temperature, wind, what? Flux measurements in general? At what time resolution?

Response: the last sentence has been modified: “To investigate seasonal patterns, further studies measuring CH_4 fluxes using micrometeorological methods at this pond or other tailings ponds during other times of the year are recommended.”