Reply to Review 2

General comments

This manuscript describes a measurement campaign conducted to measure CO2 and CH4 fluxes in a polygonal tundra in northern Siberia, with a specific interest in fluxes of a merged polygonal pond. The 2-month data has been used to show that if emissions from ponds are neglected and those from the tundra surface only are used, the upscaling to landscape fluxes underestimates the landscape CO2 uptake rate. There is not much data existing from polygonal tundra or ponds in permafrost region, and therefore I think the manuscript adds an important piece in the understanding of GHG fluxes in tundra landscape. In general, I find the paper and the data important and I suggest acceptance of the paper after minor revisions. There are certain aspects in the calculations, explanations and interpretations which are not clear to me and I think the paper could be improved by clarifying these. In addition, there are inaccuracies in the text which need to be revised before acceptance of the paper.

I suggest using overall an alternative expression for "decrease the landscape carbon sink". In my mind the sink does not decrease. I think a more descriptive expression could be e.g. "not accounting for the pond fluxes results in an overestimation of the tundra surface CO2 uptake" or similar. I understand however that in the previous literature this might have been a typical way to express the phenomenon, but it is not too late to change this.

> *Thank you for your kind words and thorough review. Please find our point-by-point replies below. Regarding the „decrease the landscape carbon sink", we agree that a different phrase might be more appropriate and changed all relevant places in the manuscript along the lines of ‚not accounting for pond emissions leads to an overestimation". This also includes the title. We propose a new title:*
> **Not accounting for thermokarst ponds leads to overestimation of tundra carbon uptake**

More specific comments

open path ch4 analyzer has been used in the study. It is well known that relatively large "Webb-corrections" are typical for that type of analyzers. I think it would be necessary to shortly discuss the implications of these corrections and perhaps mention how large they are.

> *Thank you for these suggestions. We added a discussion of these corrections in the methods:*
> This correction depends on accurate measurements of the latent and sensible heat flux and is applied to the open-path data of the LI-7700. Especially for the LI-7700, the correction term can be larger than the flux itself, but the correction is derived from the underlying physical equations. By using *EddyPro,* which uses an up-to-date implementation of the correction, and by using well calibrated instruments, we are certain to receive accurate CH_4 flux estimations from the LI-7700.

The unit of flux rate used throughout the paper is g m-2 d-1. That's OK even though not the most typically used; however it should be mentioned in the figures what do the points in the figures represent – 30 min flux, daily flux, or something else. ALso, it is crucial to tell in each figure, if just pure measurement data has been used, or if gap-filled fluxes are included also.

This is a very valid point. We changed this in all plots. One example:

Polar plot of 30-minute observed CO_2-C flux with respect to the wind direction at the EC tower.

I think it is always useful to see a time series of the original (screened) fluxes. Or if not, it is a good habit to tell how much data was available and if there were long gaps.

We added a new figure to the supplement with a time series of CO_2 and CH_4 flux data.

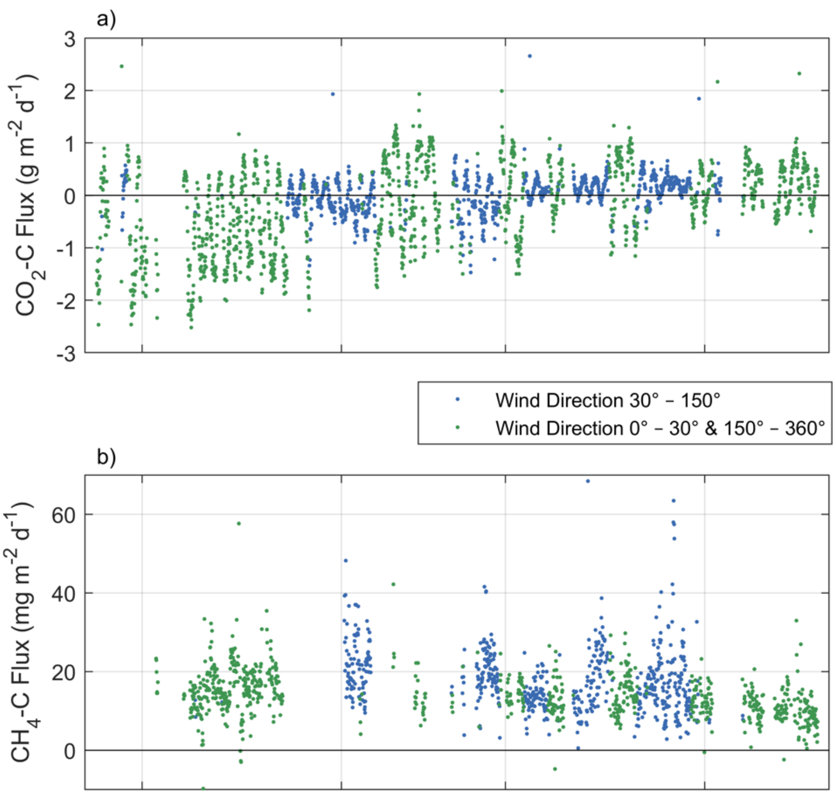


Figure A3. Time series of 30-minute observed CO_2-C flux intervals (a) and CH_4-C flux with a quality flag of 0 or 1. The blue color represents fluxes originating from the wind direction of the lake (30°–150° wind direction, mostly mixed signals from semi-terrestrial tundra and the lake surface) and the green color represents fluxes originating from all other wind directions.

line 22: "...reduce the C sink..." consider revising (see my comment above)

We agree to revise this. Please refer to our answer above to the general comments.

line 23: impact on what?

*Thank you for this comment. We added "*on the landscape-scale carbon flux*" in this sentence.*

lines 129-131: This means that there is a probability of 10% that fluxes observed at the EC tower originate from areas outside of the light gray area. Medium gray represents 50-70%, medium-dark gray 30-50%, and dark gray indicates that there is a probability of less than 30% that the observed flux originates from within the marked area.
Shouldn't it be: 10% of the flux signal originates outside the target area/fetch? Now it seems to me that you are saying that in 1/10 of the cases the whole flux data signal originates outside the area.
That cannot be the case.

That is a good suggestion. We changed the wording in the method section:
This means that it is likely that 10% of each observed flux signal originates from outside of the light gray area. Medium gray represents 50-70%, medium-dark gray 30-50%, and dark gray indicates that there is a probability of less than 30% that each observed flux signal originates from within the marked area.

chapter 2.4.3: you say you use the model to partition and gap fill the NEE data. However, I do not see a mentioning of GPP or Rtot anywhere after this chapter. Also, it is not clear for me, where you have used the gapfilled data, and where you have just averaged the accepted observations. For example in Fig. 2, is this gap filled or measured data?

We changed the first paragraph as follows:
To gap fill the net-ecosystem exchange (NEE) fluxes of CO2 we use the *bulk-NEE model* by Runkle et al. (2013). The model uses the total ecosystem respiration (TER) and the gross primary production (GPP) to gap fill NEE, our target variable.

TER and GPP are needed for gap-filling, we do not study these quantities themselves. Additionally, we add information each time we introduce a new measurement variable to explain if this is pure observation, gap filled, or a pure model result.

Line 149: "We split the datasets into a training (70%) and a validation (30%) data set to test model performance" Where do you show or discuss these test results?

Thank you for pointing this out. We added two sentences in this paragraph discussing the model performance briefly:
In 38 5-day fitting periods, we evaluate an R^2 above 0.9 between the model output and the validation set, 18 times an R^2 between 0.8 - 0.9 and six times an R^2 below 0.7. This indicates that the model works well overall.

chapter 2.4.4: modeled CO2 flux represents the vegetated tundra. So, do you use purely modeled data here, or gap-filled? If modeled, why not gap-filled? How many gaps there are in the data?

We added details to clarify the usage of modelled and gap-filled CO_2 flux, please see below.
To estimate the CO_2 flux from the merged polygonal pond (F_{pond}), we first fit the *bulk model* to data excluding fluxes from the merged polygonal pond (thus exclude fluxes >30° & <150° wind direction, as described in section 2.4.3). With this bulk model, we gap fill the CO_2 flux, and the gap-filled CO_2 flux ($F_{modeled,mix}$) represents the semi-

terrestrial tundra surrounding the EC tower, including small ponds to the north, west and south.

line 160: I'm not fully convinced why do you need the fluxes from the mixed surface to conclude something about the CO2 fluxes from the merged pond? Don't you have enough observations from that directions? Seeing the number of accepted observations and their distribution in time would help in understanding that.

We use this, admittedly slightly round-about method to extract the pond fluxes, because the merged polygonal pond is still fairly small compared to the Eddy-footprint. From Fig. 1 we learn that sometimes as much as 10% of the observed EC measurements originate from the tundra behind the merged polygonal pond, thus we have a mixed signal from all directions, also from the direction of the pond. This is why we use the bulk model to extract the pond CO_2 fluxes.
Since you already suggested that we show the accepted observations (without gap-filling), we added information about the wind sectors to these fluxes. This new figure is appended to this document and also visible in the answer of the third comment.
Finally, we tried to clarify the above argumentation by rewriting the paragraph:
Due to the strong heterogeneity of the landscape and the relatively small size of the merged polygonal pond compared to the EC footprint we measure a mixed signal from all wind directions. In other words, each flux measured with the EC method contains information from different land cover types. However, we want to extract fluxes from ponds and semi-terrestrial tundra to analyze the influence of ponds on a polygonal tundra landscape. Since we are interested in average tundra fluxes, we combine the landcover classes dry tundra, wet tundra, and overgrown water under the term *semi-terrestrial tundra*. In this way we can compare two landcover classes, semi-terrestrial tundra and the open water from thermokarst ponds.

line 161: should be >30 & < 150, right?

Yes, correct, we changed this.

line 165: "Thus, we can calculate the observed CO2flux..." this formulation sounds weird (why do you need to calculate the observed flux?), please consider revising, perhaps replace with "express" or something

Thank you for this helpful suggestion. We improved the writing as followed:
Thus, we postulate that the observed CO2 flux ($F_{obs,mix}$, not gap-filled) is the sum of the individual land cover type fluxes ($F_{modeled,mix}$ and F_{pond}) each multiplied with their weighted footprint fraction (a_{tundra} and a_{pond}), …

line 170: "To improve data quality, we exclude 30-min flux intervals of Fpond when apond<50%." Now there seems to be a contradiction: in lines 162-163 you state that Fmodeled, mix includes only data with <30% of weighted footprint fraction of open water (apond?). But now you say that you exclude all Fpond values with apond < 50%. Perhaps this needs clarification.

Thank you for this helpful comment. We see ourselves that this is contradictory. The sentence in brackets in line 162-163 was added after a suggestion of the co-authors without re-evaluation the effect on the following text, and 50% is the correct number. The text in brackets in line 162-163 does not necessarily improve the understanding of *the text. Therefore, we clarified that $F_{modelled,mix}$ is modelled and gap-filled and deleted the text in the brackets. The paragraph around line 162 now reads as follow:*

To estimate the CO_2 flux from the merged polygonal pond (F_{pond}), we first fit the *bulk model* to data excluding fluxes from the direction of the merged polygonal pond (thus exclude fluxes >30° & <150° wind direction, as described in section 2.4.3). With this bulk model, we gap fill the CO_2 flux, and the gap-filled CO_2 flux ($F_{modeled,mix}$) represents the semi-terrestrial tundra surrounding the EC tower including small ponds to the north, west and south.

lines 222-232: I have difficulties to follow the logic in this text. The chapter starts by stating that "To evaluate whether the differences in medians between the four wind sectors are significant, we apply a permutation test". Then fluxes are randomly assigned to one of two groups (why two? Ok, this comes evident when one looks at the appendix figures. But not from the text).

What is unclear to me is that how can you conclude from the test explained here and illustrated in Appendix figures that "no meteorological parameter acted as a driver for the high CH4 emission"?

Thank you for this comment. We can understand the difficulties to follow the logic here. From prior studies in the study area, we expected that of all meteorological variables high wind speed or temperature would be most likely to impact the methane emissions. So, we performed the permutation test using these two variables. To clarify this, we changed "we find no meteorological parameters acting as a driver for the high CH_4 emission" *to* "we find neither high wind speed nor high temperature acting as a driver for the high CH_4 emission"

For other available meteorological parameters, like air pressure, we tested for correlation with methane emissions, with no success.

Then, the CH4/CO2 ratios explained on lines 233-242: what is the conclusion from that analysis? I do not find any discussion about that.

You are right. This was again caused by a small misunderstanding with a coauthor. We shortened the paragraph in the results section as follow:

The ratio of CO_2C to CH_4 emissions at night (PAR < 20 µmol m^{-2} s^{-1}) has a value of CH_4/ CO_2= $0.060_{0.049}^{0.076}$ when including fluxes with an open-water weighted footprint fraction of more than 60\%, whereas the ratio amounts to (CH_4/ CO_2=$0.020_{0.015}^{0.024}$, $Median_{25\%\ Percentile}^{75\%\ Percentile}$ when including fluxes with an open-water weighted footprint fraction of less than 20%.

These results are then included in the discussion in line 334:

When looking at the night-time emissions, we find that per gram CO_2-C 0.06 g CH_4-C are emitted from ponds, and only 0.02 g CH_4-C from the semi-terrestrial tundra. This underlines again, that especially when considering thermokarst ponds, CH_4emissions are of high interest.

Lines 271-272: "Our approach of combining a footprint model with a land cover classification to extract fluxes from different land cover classes allows us to determine the pond CO2flux." This sentence is in the core of all my difficulties in understanding what has

actually been done. Didn't you use the direct pond fluxes (from sector 60-120) to infer the pond CO2 flux?

At least this is what you mention on line 207, and in the table 1. And (in my understanding so far) you used the footprint model approach to estimate the flux from "tundra" (or semi-terrestrial tundra, vegetated tundra; are these same? If yes, please ease the reader's pain and use uniform expressions here. If not, please explain more clearly what's the difference.

Thank you for this helpful comment. We have included a new method section on CH4 flux in response to the other reviewer to clarify the difference between the analysis of CO_2 and CH_4 flux:

CH_4 flux partitioning

Since we do not have a simple gap-filling model at hand for CH_4 emissions from the tundra, and since CH_4 emissions are much more variable than CO_2 emissions, we treat CH_4 differently. Instead of extracting the fluxes from the landcover types, we focus on wind sectors. We divide fluxes in wind sector:

- tundra: At least half of the footprint consists of dry tundra, and the wind direction is larger than 170°.
- $shore_{50°}$: Less than 40% of the footprint consists of dry tundra and water contributed to the footprint with at least 30%. The wind direction lies between 30° and 65°.
- pond: At least half of the footprint consists of open water and the wind direction lies between 65° and 110°.
- $shore_{120°}$: Less than 40% of the footprint consists of dry tundra and water contributed to the footprint with at least 30%. The wind direction lies between 110° and 130°.

CH_4 permutation test

To evaluate whether the differences in medians between the four wind sectors are significant, we apply a permutation test (Edgington and Onghena, 2007). In this test, we randomly assign each 30-min flux signal to one of two groups, calculate the median of both groups and their difference (with four groups it amounts to six tests in total). After repeating this step 10000 times, we plot the resulting differences in medians in a histogram and perform a one-sample t-test to evaluate whether the observed difference in medians differs significantly ($p<0.01$) from the randomly generated differences.

We also clarified in (the former) line 207 that the pond CO_2 flux was estimated using equation 4 from the method section. We hope that this makes our approach clearer. To improve readability, we added a sentence in the section of "Open-water CO_2 flux" that defines the semi-terrestrial tundra, and we now always only use this term, and not 'vegetated tundra'.

Since we are interested in average tundra fluxes, we combine the landcover classes dry tundra, wet tundra, and overgrown water under the term *semi-terrestrial tundra*. In this way we can compare two landcover classes, semi-terrestrial tundra and the open water from thermokarst ponds.

Chapter 4.2: The observation of the CH4 spike in the shore120 is interesting, and the fact that it remains unexplained, is pity but not unexceptional in flux studies! It is also

somewhat convincing how much effort you have had to explore the reasons for the higher emission

Thank you very much.

line 340: a somewhat similar approach has been used also earlier, see e.g. https://bg.copernicus.org/articles/16/255/2019/

Thank you for this comment. In the method section we tried to embed our approach better into the existing literature and we now refer to this work and the work by Rößger et al. (2019a); Rößger et al. (2019b) by adding a new sentence:

Similar approaches of analyzing heterogeneous eddy covariance fluxes in the arctic environment have been conducted for CO_2 and CH_4 (e.g. *Rößger et al. (2019a); Rößger et al. (2019b)*b; Tuovinen et al. (2019)). Rößger et al. 2019 a,b extracted CO_2 and CH_4 fluxes from two different land cover classes on a floodplain, and Tuovinen et al. (2019) separated CH_4 fluxes from nine different land cover classes, including water, and combined them into four source classes. All three studies have in common that they differentiate fluxes from different vegetation types, however our method is dedicated to differentiating between fluxes from tundra and water.

Figures 2 and 5 (which are nice and indicative figures overall!): please indicate if the fluxes consist of purely measured values, purely gap-filled, or both. If just measured values are shown, how are the mean values (in red) used in the study? If there are missing values during the day, the mean does not represent the true daily NEE. Is the mean (red value) a mean of all fluxes from that direction during the 2-month period?

Thank you for this suggestion. We added the information about observed or modelled data in all plots and included the following words in Figure 2 and 5 to clarify that the 5° average was taken for the whole 2-months period:

during the 2-months observation period

Figure 3: is each dot a 30-min flux? Please explain it. “Flux intervals at night time”? Why interval, aren’t these just fluxes?

Thank you for this comment. We think that confusion arises, because we used the term “flux interval” for 30-minute fluxes. In our view one 30-minute flux can be also defined as a flux interval. For clarification, we added the term “30-minute” in front of “flux” and removed the “interval”. We also edited all other parts in the manuscript in this way.

Fig 4: please explain how the violin plot should be interpreted

Thank you for this suggestion. We added to following text at the end of Fig. 4 and 6:

A violin plot shows the distribution of measurements along the y-axis - the width of the curves expresses the density of data points at each y-value.

Fig A2: what is the red line?

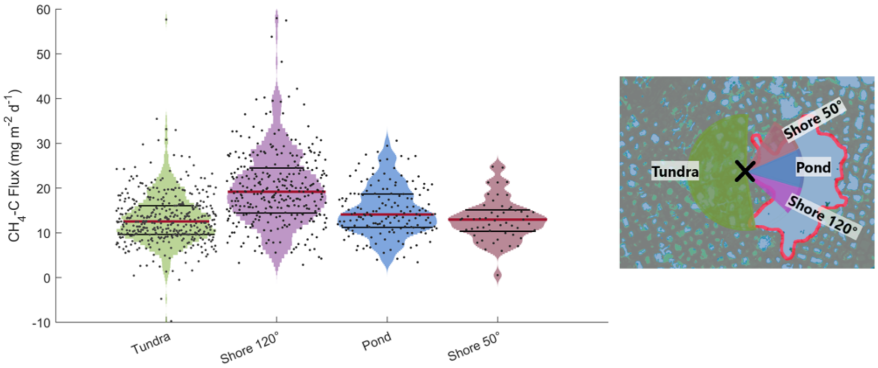
We added to following in the figure caption:

The observed differences in medians between the different wind direction classes are shown in red vertical bars in each plot.

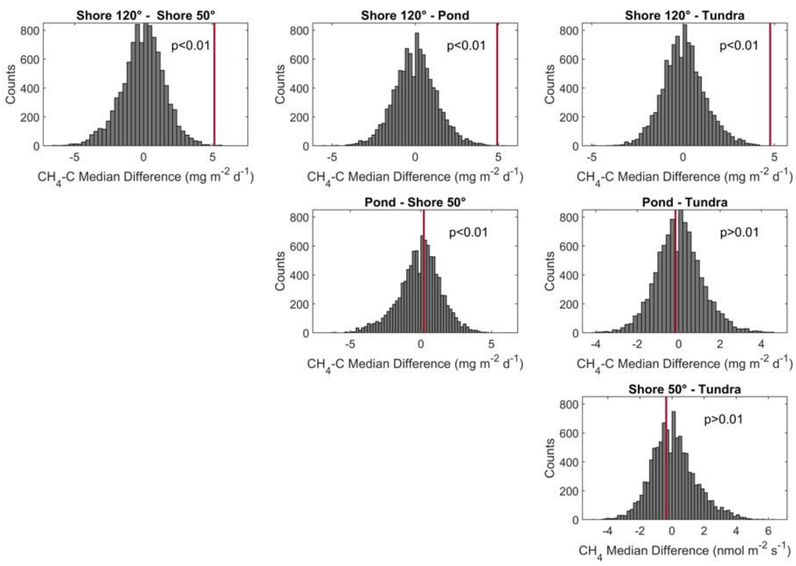
Figs A2 & A3 and line 310: there are no b’s or c’s in Fig. 6

Thank you for noticing. These letters were included in a previous version of the manuscript. We deleted the letters.

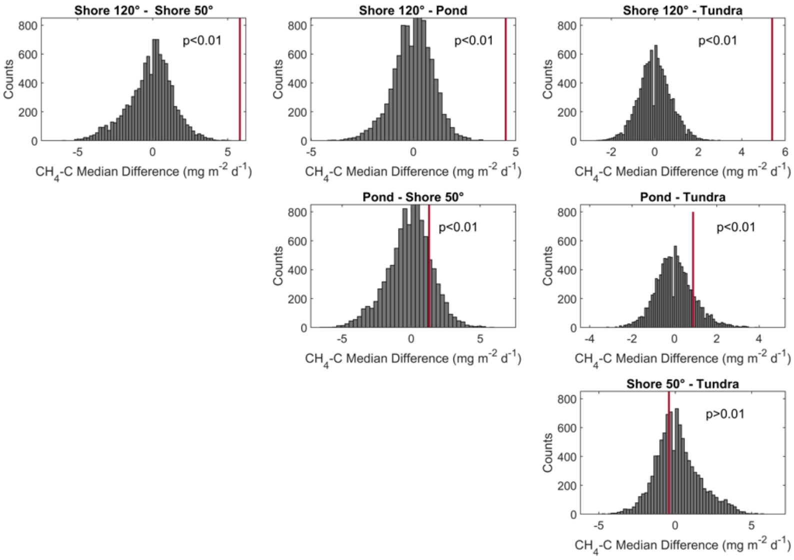
Additionally, while working on the suggestions by the reviewers, we found a small error in the program code. We corrected for this error and updated the figures 6, A4, and A5 and updated the numbers on CH4 flux in table 1 and section 3.3. The error had only a small impact and had only a noticeable effect on the statistical test of figure A4, center graph. Now, the difference in medians between *pond* and *shore*$_{50°}$ is significant and we described this in the discussion. We also updated the resulting numbers on *pond* CH4 flux (median *pond* CH_4-C flux before correction 13.38 mg $m^{-2}d^{-1}$, after correction 13.90 mg $m^{-2}d^{-1}$). We also had to update the standard deviation of CO2 flux at the beginning of the result section. However, all the conclusions we draw remain unchanged.



Updated Figure 6. Violin plots of observed CH_4 emissions at the EC tower separated into four different wind direction classes. A violin plot shows the distribution of measurements along the y-axis - the width of the curves expresses the density of data points at each y-value. Medians of CH_4 emission distributions are shown as red lines, and 75th & 25th percentile are shown as black lines. On the right, the wind sectors with the eddy covariance tower in the center (black cross) are shown.



Updated Figure A4. Histogram of permutation tests between the medians of CH_4 emissions from different wind direction classes in figure 6. All medians from flux observations during moderate wind speed conditions. The observed differences in medians between the different wind direction classes are shown in red vertical bars in each plot.



Updated Figure A5. Histogram of permutation tests between the medians of CH_4 emissions from different wind direction classes in figure 6. All medians from flux observations during moderate air temperature conditions. The observed differences in medians between the different wind direction classes are shown in red vertical bars in each plot.

Literature

Edgington, E. and Onghena, P., 2007. Randomization tests. CRC press.

Rößger, N., Wille, C., Holl, D., Göckede, M. and Kutzbach, L., 2019a. Scaling and balancing carbon dioxide fluxes in a heterogeneous tundra ecosystem of the Lena River Delta. Biogeosciences, 16(13): 2591-2615.

Rößger, N., Wille, C., Veh, G., Boike, J. and Kutzbach, L., 2019b. Scaling and balancing methane fluxes in a heterogeneous tundra ecosystem of the Lena River Delta. Agricultural and Forest Meteorology, 266-267: 243-255.

Runkle, B.R.K., Sachs, T., Wille, C., Pfeiffer, E.M. and Kutzbach, L., 2013. Bulk partitioning the growing season net ecosystem exchange of CO₂ in Siberian tundra reveals the seasonality of its carbon sequestration strength. Biogeosciences, 10(3): 1337-1349.

Tuovinen, J.P., Aurela, M., Hatakka, J., Räsänen, A., Virtanen, T., Mikola, J., Ivakhov, V., Kondratyev, V. and Laurila, T., 2019. Interpreting eddy covariance data from heterogeneous Siberian tundra: land-cover-specific methane fluxes and spatial representativeness. Biogeosciences, 16(2): 255-274.