Answer to Reviewer #1

Thank you very much for your comments and support for the publication of our manuscript. Below we address one by one the comments made during this review. All answers are in blue font.

Specific comments

In my opinion, the comparison between observed and simulated methane emissions would however benefit from using an upscaling approach to avoid issues arising from the mismatch of scales. This was done for the chamber measurements, but it remains unclear how representative the flux tower footprint is of the entire grid cell. Comparing flux measurements from a single location to the entire grid cell is only meaningful if the grid cell is characterized by spatially homogeneous methane emissions. This is only rarely the case for such high-latitude landscapes (e.g., Sachs et al., 2010; Parmentier et al., 2011; Helbig et al., 2017).

We agree that the comparison between model methane fluxes and those from observations, specifically from eddy covariance, is a challenge. In our manuscript, we use a scaling factor for the chamber data by considering chamber measurements that were done under exclusively wet and under exclusively dry summer conditions. We then make use of the total fraction of inundated areas in the model grid cell (IF) modeled with the TOPMODEL approach to scale the total chamber fluxes. This scaling approach takes into consideration that the model methane fluxes represent the emissions from only the portion of the grid cell that is inundated, i.e. with water at or above the soils surface.

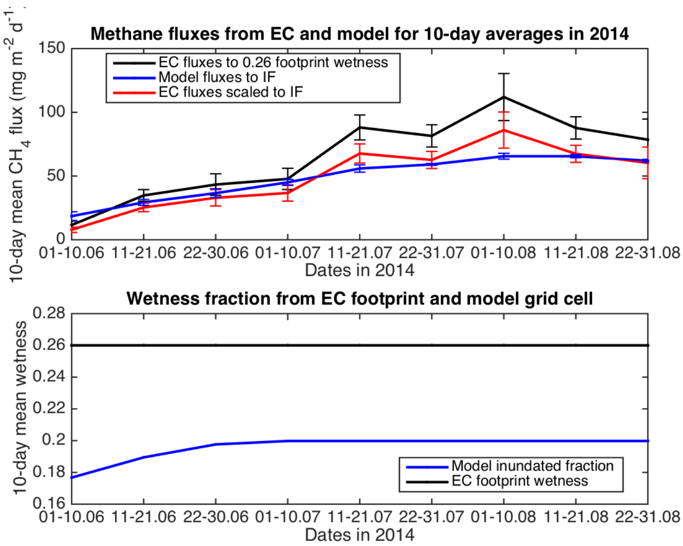
In the case of the eddy covariance fluxes, following the concerns of the reviewer, we re-evaluated our approach for this comparison. In the revised version of this manuscript we include now a thorough analysis of the footprint area of the eddy covariance fluxes as part of a new Appendix B on "Details on in-situ flux observations". This appendix also includes details on the eddy covariance flux data uncertainty assessment and more detailed results on the chamber measurements, as requested below also by the reviewer. This appendix will be part of the revised manuscript and is attached at the end of this response.

In this new appendix, we analyze the type of vegetation and its coverage in the footprint area of the EC tower, from remote sensing images as a metric to identify wet and dry areas. Areas with dominant cotton grasses, specifically *Eriophorum angustifolium* in our study area, are indicators of predominant wet soils, while tussocks, specifically *Carex appendiculata* in our study area, and shrubs are indicators of predominant dry soil conditions. It is important noting that *C. appendiculata*, can be also found in wet areas, but is predominant in dry areas.

For the model, the vegetation distribution per grid cell is too coarse to consider this metric similar as that for the remote sensing data in the EC footprint area, however the total abundance of C3 grasses in the grid cell A is 33.3 % as given for the model (with the rest of the grid cell dominated by deciduous shrubs and extra tropical evergreen trees), but there is no discrimination between cotton grasses and tussocks.

The footprint of the eddy covariance tower in the Chersky floodplain covers an approximate area of 400 m x 400 m, similar to that one depicted in Fig. 1 of Kittler et al. 2016 (cited in discussion ms) (see new Appendix B at the end of this response for footprint area for the EC tower used in this manuscript). The remote sensing analysis revealed that cotton grasses are present in about 26 % of the footprint area, which would translate into the same portion of the footprint area as fully wet zones during the "wet months": after spring melt in June and until August when most annual precipitation in the region takes place, covering most of the growing season. As will be shown below in this response, CH_4 fluxes measured by chambers (footprint of 60 cm x 60 cm) revealed that during the growing season in dry soil areas of the Chersky floodplain that are characterized by a water table below the surface, the emission of methane during the growing season is negligible with even some atm. CH_4 uptake by soil (i.e. negative CH_4 flux rates) (data shown in new Appendix B). Under this consideration, and as confirmed recently by Helbig et al., 2017, the majority of the CH_4 fluxes measured by the EC tower would represent fluxes from fraction of wetland in the footprint area, i.e. 26 %.

In case of the model grid cell where the location of the EC tower falls (grid cell A in Fig. 1 of the discussion ms), the IF for June-July-August during 2014 shows growing inundation values from 17.7 % to 19.9 % (for 10-day mean values for those three months) representing the percentage of total wet areas in the grid cell area. These values are slightly smaller than the 26 % wetness area in the EC footprint, and denote the area of the grid cell where the model methane emissions take place (i.e., no emissions in dry areas, in agreement to the chamber measurements).
With this basis and to make a closer comparison between EC flux measurements and model data for the growing season months, we scaled linearly the 10-day mean EC methane fluxes to the IF from the model, and calculated the standard deviation of the 10-day mean. In the next figure, we show: TOP panel, the original 10-day mean EC methane flux measurements that would represent the emissions of a 26 % wet area between June and August 2014 (black line), the 10-day mean EC methane fluxes scaled to the 10-day mean IF from the model for the same period of time (red line) and 10-day mean model methane emissions for grid cell A, which imply emissions from the IF from the model (blue line). Error bars in all lines are one standard deviation of the 10-day mean flux values. The BOTTOM panel shows the 10-day mean IF from the model used to scale the EC fluxes (blue line), and the constant wetness percentage of the footprint area calculated from the vegetation coverage remote sensing images (i.e., 26 %).



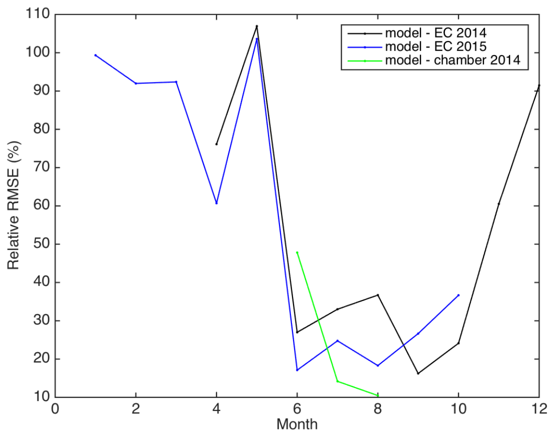
We observe that the scaled EC methane fluxes decreased as a lower IF is considered within the footprint, and those new calculated fluxes become closer to those from the model, and in most cases the latter fall within the 10-day standard deviation of the EC fluxes.
Unfortunately, it is not possible to obtain a temporal varying wetness area for the EC footprint all year, based on our approach of only considering the vegetation cover, thus wouldn't be appropriate to scale all of the EC fluxes for 2014 and 2015 to the IF from the model without any reference for spring and winter wet footprint areas. However, from this analysis we learn that: 1) considering the vegetation cover as indicator of soil wetness, the EC footprint area holds a very similar area to that of the model grid cell through which the majority of the methane is emitted to the atmosphere and 2) the net offsets between methane flux model and EC data can largely be attributed to differences in wetness levels.
Summarizing, we assume that for both the model grid cell and the eddy covariance footprint, methane emissions are not spatially homogeneous, but bound to the distribution of wet (inundated) areas. Accordingly, a meaningful agreement between model and observations can only be obtained if two factors are fulfilled: (i) the fraction of wet surfaces agrees between both data sets, and (ii) the flux rates from wet surfaces agree between both datasets. Through correcting the offsets in inundated fraction, we could demonstrate that the flux rates between model and eddy covariance observations agree very well, emphasizing the sound setup of the

model algorithms and parameter settings. We will add the analysis presented here into the new Appendix B to complement the discussion on scaling fluxes for comparison between EC and model data.

The authors should also address how representative the location of tower and chamber flux measurements is of the entire grid-cell. The authors estimate the fraction of inundated land for the grid-cell and demonstrate how this fraction is an important predictor for methane emissions. The same should apply for flux tower measurements where the fraction of wetlands is tightly coupled to the magnitude of methane emissions (see for example Helbig et al., 2017). How would the wetland fraction at the grid cell-scale compare to the same fraction at a smaller scale at the study sites?
We approached this comment with the answer above. By evaluating the vegetation cover types within the footprint area of the EC tower, we identified the wet areas and assume that the methane fluxes measured with this tower represent the emissions from the wetlands within the footprint. Equivalent to the grid cell area, the inundated fractions determined with the TOPMODEL approach, represent the areas where methane is emitted at grid cell scale. Those are comparable and to show this, a scaling exercise for growing season methane emissions in 2014 was presented above.

The authors report "comparable" (line 30) methane emissions when comparing model and measurements. The analysis could be much stronger if the authors give a quantitative measure for the performance (e.g, Root Mean Square Error or any other suitable metric).
As suggested by the reviewer, we include now in the revised ms the relative RMSE calculation in percentage (e.g. RMSE / mean(CH4_obs) * 100) between model and flux measurements from Eddy Covariance (for 2014 and 2015) and chambers (only for the available three months in 2014). We calculated this error on a monthly basis, using the daily resolution fluxes. Results are shown in the table and figure below.

Month	Rel. RMSE (%) (model – EC) 2014	Rel. RMSE (%) (model – chambers) 2014	Rel. RMSE (%) (model – EC) 2015
Jan	-	-	99.3
Feb	-	-	91.9
Mar	-	-	92.4
Apr	76.1	-	60.7
May	106.9	-	103.6
Jun	26.9	47.8	17.1
Jul	33.0	14.2	24.7
Aug	36.7	10.5	18.3
Sep	16.2	-	26.6
Oct	24.1	-	36.6
Nov	60.5	-	-
Dec	91.4	-	-



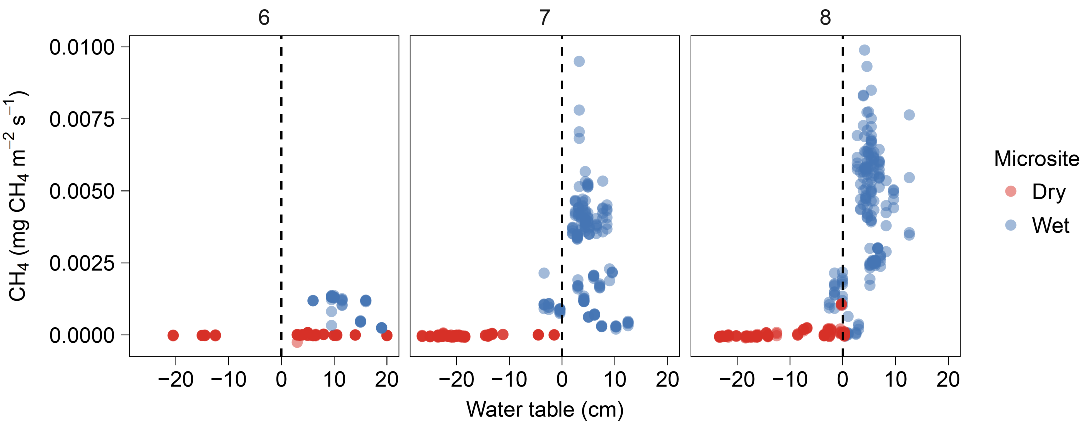
The relative RMSE results show the relative variation between the model and the observations. A larger variation is observed in the first five and last month of the year (winter and spring) between model and measured EC fluxes, while the lowest variations are observed during the growing season and autumn (June to October). The summer variation is larger in 2014 between model and EC data and lowest in July and August between the model and chamber measurements in 2014. This information will be included in the revised MS to quantitatively support the evaluation of the model results.

The authors state that the aim of the work is to "improve our understanding". However, in my opinion, the manuscript mainly focuses on improvements in methane modeling and an evaluation of the performance of a revised methane model. The authors may consider reframing their research objectives and focus results and discussion on the specific research questions.
The reviewer is correct that the stated aim is not reflecting the bottom line of our manuscript. Following this suggestion, we reframed the aim to be clearer and now it reads: "The aim of this work is to analyze the performance of an improved process-based methane model, designed for Arctic tundra and wetlands underlain by permafrost, when applied to a regional domain in Northeast Siberia. Our intention is to evaluate the potential of a refined process-based methane model as a proof of concept, for its application to a larger than site level scales. For this, year-round CH_4 emissions are modeled and differentiated among distinct pathways: plant-mediated, ebullition, and diffusion." We also focus the discussion towards this aim in the revised ms.

Large areas in northern Siberia are covered by polygonal tundra. The distinct microtopography of these landscapes has important implications for surface hydrology and thus also surface inundation (see Cresto-Aleina et al., 2013; Helbig et al., 2013; Liljedahl et al., 2016). I was wondering if such polygonal tundra covers a significant proportion of the study area?
And if yes, what would be the consequences of distinct microtopography on the performance of the TOPMODEL and on the simulated methane emissions. Using a mean water table for methane modelling in such heterogeneous landscapes can lead to significant underestimation of methane emissions (Cresto-Aleina et al., 2016).
The reviewer is right that a good portion of the Siberian tundra is characterized as polygonal tundra. However, our area of study does not contain these particular micro-topographic structures since it is mostly located in a floodplain that naturally becomes inundated at the end of the melt season (spring). Towards summer, most of the water recedes to streams and to the Kolyma River and nearby tributaries only to lead to a typical wetland landscape. Still, some polygonal structures are present, but they are not a dominant feature of the landscape within our model domain, as opposed to e.g. the Lena River delta. Therefore, the application of TOPMODEL in the Chersky floodplain is suitable and there is no need to consider polygonal structures.

With the TOPMODEL approach, the authors can distinguish between inundated and non-inundated land. However, many peatlands are characterized by a water table just below the peat surface and are thus not inundated. Nevertheless, they can emit large amounts of methane, which would be neglected in the current modeling approach.
We are aware of the limitations on the use of TOPMODEL in those particular cases where the water table is located below the surface and those were discussed briefly in the discussion manuscript. The study of Kwon et al. (2016) (cited in the discussion ms) reported the flux chamber measurements in the same Chersky floodplain site subject to our study. The authors reported CH_4 fluxes measured in plots where the water table was 10 cm below the surface and found negligible contribution of CH_4 from these soils. Specifically, areas with water table 5 cm below the surface showed net CH_4 emissions but flux rates were not as high as in areas with standing water (see figure below). Also, as shown in the new Appendix B attached at the end of this response, chamber flux measurements of CH_4 in dry soils with water tables ca. 10

cm below the surface show none or negligible methane emissions to the atmosphere in this area of study. Taking into account these findings, in our model configuration the fact that no methane emissions take place in dry soils, would not pose a constraint to the total modeled methane fluxes per grid cell in this area of study, however, the role of methane oxidation could be better evaluated.



The figure above shows results of CH_4 chamber fluxes (mg CH_4 m^{-2} s^{-1}) measured from June to August (numbers at the top of the figure indicate the month of the year: 6 is June, 7 is July and 8 is August) in 2014 for the dry and wet plots and their corresponding water table (in x-axis). “Dry” plots have mostly water tables at or below the surface during July and August with mostly uptake of CH_4 from the atmosphere (on average 3 mg CH_4 m^{-2} d^{-1}), whereas the wet plots characterized by water tables located above the surface, showed average emissions of 332 mg CH_4 m^{-2} d^{-1} over the same period of time.
Despite this agreement, we are aware that the low CH4 uptake in dry areas might not apply to other tundra areas, e.g. in Zona et al., (2016) in the Alaskan tundra the highest fall and winter CH_4 fluxes were observed in upland tundra sites characterized by having a water table below the surface during summer. In future studies, our model scheme should also be tested in other areas such the Alaskan tundra to assess and improve further the model configuration especially in the TOPMODEL scheme.

At the same time, lakes (i.e., inundated land) may be characterized by lower methane emissions than these peatlands due to a lack of fresh organic carbon input. What are the implications of this for the modeling performance? The authors may consider discussing this shortcoming.
This is an interesting idea and we agree with the reviewer that a comparison of lake and peatland model results would be an ideal evaluation of our methane scheme using extreme cases of water table depth. However, we do not see the possibility to perform such study, as explained below.
In our model configuration, the production of methane is considered to take place in mineral soils and does not include peatlands as definition: the layer of soil with > 30 cm of organic rich material (peat) accumulation. A mask containing the distribution of peatlands should be needed to introduce this feature. In addition, as carbon decomposition slows down in permanently anoxic areas of the soil column, the prescribed mask of peatlands should contain added soil C in order to describe deep peat layers characterized by a slow decomposition timescale. These steps are currently been taken for the global context with the JSBACH model and are still pending work for high horizontal-resolution domains such as the regional one presented in this work.
The scheme to model wetland areas using the TOPMODEL approach considers the topographic profile, which is provided as a prescribed compound topographic index in the model domain, and methane emissions take place in areas where the water table is located at or above the soil surface. In this context, the model does not explicitly simulate the location

of "lakes" (inland open water bodies) but rather a dynamic change in the horizontal distribution and accumulation of water at or above the surface, which in turn may consider implicitly inland water bodies at different scales: lakes, wetlands, ponds, etc. With this model, it is not possible to discriminate at this coarse resolution, the type of water bodies, but rather it provides an average portion of the grid cell area where inundation can take place, and only the methane production and ultimately emissions, are linked to the carbon content and environmental conditions of the soil. If by definition there is no consideration of peatlands in our model, in the end all goes down to the available organic carbon in the soil for the production of methane. As requested by the reviewer, we will discuss this shortcoming in the revised manuscript.

In the current manuscript, the authors "decreased or increased [the parameters] by a fixed value" (line 343). Could the authors use a Monte-Carlo approach instead to assess the parameter sensitivity?
The purpose of the parameter permutation is to know, to which parameter the model is most sensitive, as this identifies which parameter need to be better constrained to reduce model uncertainty. The purpose of a Monte-Carlo approach is the identification of the uncertainty of a model given a known probability distribution function of parameter values for a combination of parameters. MC approaches are not primarily designed to identify model sensitivities to specific parameters, even though some approaches such as LHS allow interpreting MC approaches in terms of model sensitivities; however only at very high computational costs. One-at-a-time schemes (OAT) as the one applied here directly target the model sensitivity, are computational cost efficient, and are deemed fully sufficient for the purpose (Saltelli et al. 2000). The identification of compensating effects between parameters or non-linear effects, which would require an MC approach area, is beyond the scope of this paper. We therefore consider an MC approach to assessing model sensitivity as unnecessary.

The authors mention "reported values in the literature". Could they specifically discuss/show the observational constraints on the individual parameters?
We thank the reviewer for this comment. We improved the description of our selection of parameters for the sensitivity study in the revised ms. The selected parameters are those that are prescribed in the model and are considered uncertain. Specifically, the selected values for ϕ (snow porosity) and *f*CH4anox (fraction of anoxic decomposed carbon that becomes methane) were those kept within ranges of values previously discussed in the published literature, whereas for the other four selected parameters (see below) we chose a range of values around the defined values for the control simulation.
Thus, the selected parameters are characterized by at least one of the two criteria: 1) it is a parameter with large uncertainty because it is not provided in current published literature or its values are still controversial as reported in published literature, and 2) it is possible to test a range of values based on reported values in literature. The last criterion is only true for two of the selected parameters (ϕ and *f*CH4anox) as mentioned above. As given in the discussion ms, the six selected parameters for our sensitivity studies are:
In the TOPMODEL scheme:

1) χmin_cti, minimum compound topographic index threshold value. This parameter fulfills criterion 1 since it is a model parameter that is exclusively part of the TOPMODEL scheme, therefore there is no literature reference and rather is a given value that has to be adjusted.

In the plant-mediated transport scheme:

2) *d*r, root diameter. This is a highly uncertain value in literature with only few reported values. Few studies have reported the diameter of vascular plants in boreal ecosystems. In Wania et al., 2010 (cited in discussion ms and after Schimmel, 1995) the authors report a diameter for *Eriphorum angustifolium* of 3.95 mm, while for *Carex aquatilis* a value of 3.8 mm. Chapin and Slack (1979) reported a diameter for *Eriophorum vaginatum* of 0.8 mm, while Wang et al., 2016 reported a value of 1 mm for the same species. For our model set up, we use a value of 2 mm in the control run

considering an average value between those reported in the literature. For the sensitivity study, we selected higher root diameters experiments: 5 and 8 mm.

3) *Rfr*, principal fraction of the pore-free soil volume occupied by roots. This is also a highly uncertain value that is not reported in literature; therefore, we assume in our control experiment a fraction of 40 % (i.e., in a certain volume of soil, 40 % is occupied by plants roots). For the sensitivity studies, we decreased and increased this reference value by 50 % of the control value, i.e., 20 % and 60 % respectively.

In the diffusion of gas through snow:

4) h_{snow}, snow depth threshold. The studies of Pirk et al., (2016) and Smagin and Shnyrev (2015) (both cited in the discussion ms) measured CH_4 emissions through snowpacks under different conditions. These studies evidence the transport of gas through snow layers as thick as 1.4 m. However, regarding the thinner snowpack the authors only show results from layers 10 cm thin. For our purpose, the lower limit of snow thickness is simply a model metric that allows us to differentiate between emissions in the presence or absence of snow. We selected thinner snow layers to test the model response, and the changes on this threshold thus mainly determine the timing of the emissions, which ultimately influences the magnitude of the total emissions through snow by allowing an earlier or later release of gas trapped in the soil.

5) ϕ, snow porosity. This parameter has been previously reported in literature and is derived from snow density measurements, which ultimately controls the amount of gas that can be diffused through the snow layer. This was discussed in the ms. Based on observations, Pirk et al., (2016) measured methane emissions through snow with densities that ranged between ca. 250 kg m^{-3} (at the surface of the snowpack) to 420 kg m^{-3} (at about 80 cm depth of the 1.4 m snowpack). According to our model results, the snow depths in the model domain did not exceed 30 cm during the peak of the snow accumulation (shown in Figure S4c of the discussion ms), thus is unlikely to find dense snowpacks. We chose a maximum density of 330 kg m^{-3} that corresponds to a porosity of 0.64 as our control value and tested for the sensitivity experiments less dense snowpacks with increasing porosities of 0.71 (for a density of 263 kg m^{-3} characteristic of aged snow) and 0.86 (for a density of 128 kg m^{-3} for fresh snow).

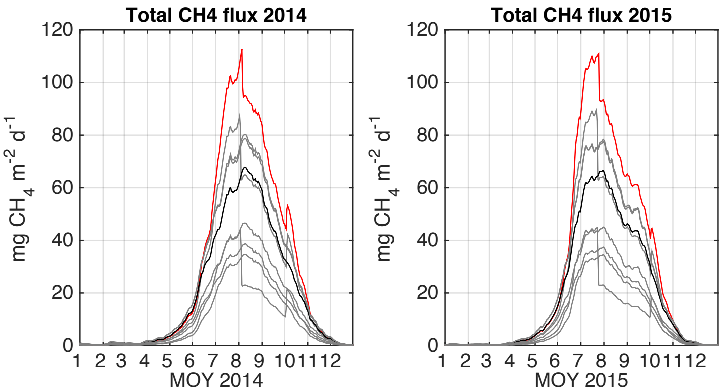
In the overall methane module:

6) *f*CH4anox or the fraction of anoxic decomposed carbon that becomes methane. This is a highly uncertain parameter in literature with some reported values in literature. In the discussion ms, we thoroughly discussed it (Lines 882 to 914), therefore we refrain to include here this discussion. However, we summarize by arguing that despite there are some values reported in literature, these are still uncertain and in our sensitivity tests we chose those values that have been reported and are characteristic of specific field conditions.

In the revised ms, we will improve the description of the selected parameters, especially those that are obtained from observational constraints.

Line 406-408: Why do the authors only show one adjacent cell? What is the justification to compare a neighboring grid cell to the ground-based observations? To demonstrate the spatial heterogeneity the authors could consider using more than just two grid cells.

Thank you for this suggestion. Our intention to show a neighboring grid cell was to demonstrate the spatial heterogeneity in the model results. Showing other grid cells in the model domain indeed can complement this. We believe that the maps showing spatial variability in flux rates provide already a good overview on the overall spatial variability. This larger scale variability is a superposition of many environmental factors, the most important of these being inundation fraction and coverage fraction of C3 grasses. The closer analysis for these two cells that the reviewer refers to was mainly performed to emphasize that even moderate variations in these factors (and others, such as e.g. soil depth) can lead to systematic differences in simulated fluxes. As we see it, extending this kind of analysis also to other cells would not add to this message, but rather confuse the reader by providing too much information. We suggest, however, to extend the discussion related to the spatial heterogeneity in the modeled methane emissions by showing results of mean total methane fluxes in the eight grid cells surrounding grid cell A.



The figure on the left shows the time series of the total CH4 fluxes at daily resolution for the eight grid cells surrounding grid cell A (shown in black). One of those surrounding grid cells is grid cell B (red line).

In the tables below, the mean±std. of the total methane fluxes during summer (June, July and August) from grid cell A (given as values in black at the center cell of each table), and the surrounding eight grid cells for 2014 (left table) and 2015 (right table). Left side grid cell from the center cell, corresponds to the values for grid cell B (values in red).

June, July and August in 2014

25.4 ± -7.1	56.0 ±-12.1	23.7± -3.7
72.8±-15.9	48.6±-5.5	27.8±-4.0
57.2±-6.8	57.9±-6.9	33.4±-4.7

June, July and August in 2015

25.2±-8.3	56.9±-14.7	24.9±-4.8
75.5±-18.3	50.1±-7.3	28.2±-4.8
59.2±-8.7	59.5±-10.4	34.3±-5.7

In line 464-465, the authors mention the "parameter adjustment", but do not elaborate how exactly the parameter for the TOPMODEL was adjusted. Did the authors use an objective (cost) function to optimize this parameter?

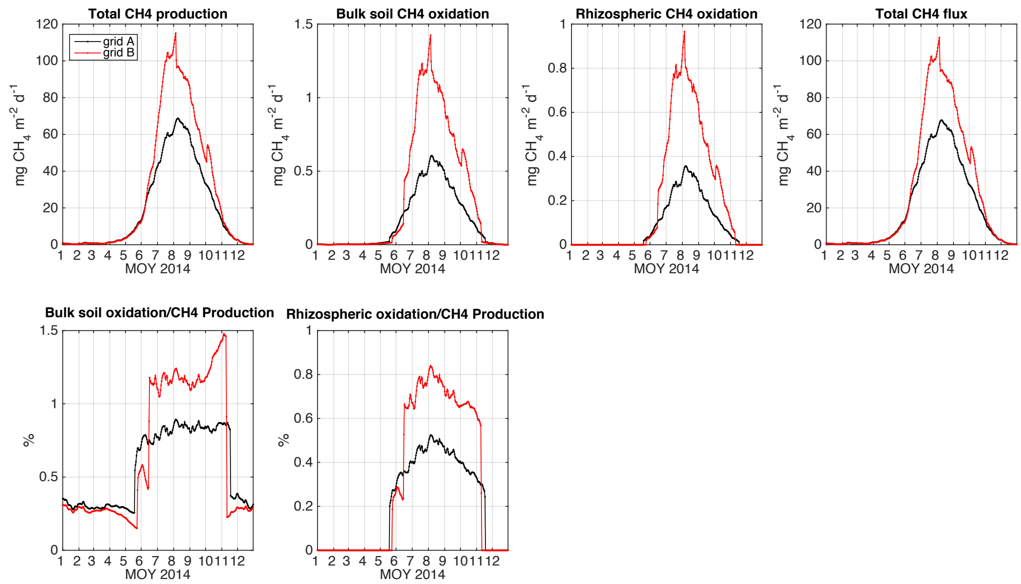
There was no optimization of these parameters based on a cost function. The parameter adjustment for the TOPMODEL was also done in the same fashion as for the sensitivity studies: by varying each of the parameters of the TOPMODEL and analyzing the response by comparing the output to the chosen remote sensing data. This model parameter adjustment can only be done in this way within the current model structure. A more sophisticated optimization of parameters falls into a model data assimilation type, which is not implemented in this model configuration and goes beyond the scope of this work.

The authors demonstrate in their sensitivity analysis that the threshold TOPMODEL parameter and "allocation-of-decomposition-to-CH4" are the most important parameters determining the magnitude of simulated methane emissions. In my opinion, the authors should strengthen these results throughout the manuscript. It appears as if their results indicate that methane emissions mainly depend on methane production dynamics (i.e., fCH4anox) and on inundation as "on-off" switch of methane emissions.

The threshold TOPMODEL parameter and the allocation of C decomposition to methane are the parameters that, under the current model configuration, settings and for the selected groups of parameters for sensitivity tests, the most influential to the simulated methane emissions. This test was aimed to identify which of the most selected uncertain parameters have the highest influence to the results and with that, identify where the model is more sensitive and where it needs further improvements and evaluations, i.e. especially in those processes where the most influential parameters play a role in the model as in this case in the hydrology and carbon decomposition.

The methane emissions in our process-based model, not only depend on the methane that is produced based on the available carbon decomposed in the soil, but also depend on the available volumetric soil pore space, moisture, soil temperature and ice content in the soil as driving processes. Indeed, once methane is available in the soil to be emitted to the atmosphere, the inundated areas simulated with the TOPMODEL approach, determine the magnitude of the emissions. Our discussion regarding the sensitivity studies is based solely on the parameters chosen for the sensitivity experiments, and those are the threshold parameters in TOPMODEL and the fraction of available carbon to be decomposed into methane. We will improve this discussion to emphasize this result in this section.

Transport pathways and methane oxidation appear to be less important (merely changing the timing of emissions). Are these modelling results supported by observations in the field? The authors may consider discussing this in more detail.
The methane transport pathways are the result of the process-based methane calculations in the model according to, among others, the changes methane and oxygen concentrations in the soil and in the soil pore space that varies in relation to the freezing and thawing soil cycles, influencing directly the methane concentration in the soil.
The timing of the emissions is linked to the changes mostly in the soil physical state and speed of transport processes by their definition, e.g. diffusion of gas in air is faster than in water, resistance to molecular gas diffusion through the exodermis of plants, all in a process-based design. The model still lacks of a proper hydrology representation that allows inundation without having to set the soil moisture to saturated conditions, and that overall has an impact in the e.g. diffusion and oxidation of the methane. In the parts of the grid cell that are not water saturated because inundation cannot take place, the methane processes are not taken into account. Thus, the still not well-represented methane processes are not less important, but are only part of the limitations of the current model configuration and these results hint to the next steps to improve the model.
As presented in the discussion manuscript, some field studies have conducted experiments to measure independently the different pathways of methane emissions into the atmosphere. Through isotopic quantification of $\delta^{13}C$, Knoblauch et al., 2016 (cited in discussion ms) measured the amount of methane emitted through plants; Kwon et al., 2016 using chambers in the Chersky floodplain, also measured the gas emitted through plants. These studies were discussed in the discussion manuscript, and we demonstrated that, in agreement to field studies, the most dominant methane transport pathway from the total annual emissions (ca. 70 – 90 %), is through vascular plants when they are present. In the case of ebullition, this is a more difficult process to measure in field studies, because of its episodic nature. Despite some studies have attempted to measure methane emitted exclusively through ebullition (Tokida et al., 2007; Jammet et al., 2015 both cited in discussion ms), for models it is difficult to evaluate this process against observations.
In the case of methane oxidation, in our model configuration the oxygen content is explicitly taken into account, enabling two process-based oxidation processes: bulk soil methane oxidation and rhizospheric methane oxidation. After methane is produced in the soil (from available decomposed carbon), the bulk soil methane oxidation can take place considering the available oxygen in the soil pore spaces. The other oxidation pathway considers the available oxygen in plants. Only part of the oxygen in the soil is available for methane oxidation, and this discrimination relates to the amount of carbon dioxide produced during heterotrophic respiration, which has a maximum value of 40 % of the total oxygen content in the soil. An additional 10 % of the available oxygen is assumed to be unavailable because it is used in other processes (e.g. respiration by microbes). This leads to only 50 % of the total oxygen in the soil to be available for CH_4 oxidation. The methane processes in the model (oxidation and emission) take place in the inundated area, and this also restricts the magnitude of the oxidation. The daily methane oxidation rates for the two oxidation pathways for grid cells A and B in 2014 are shown in the figure below.



The bulk soil CH_4 oxidation accounts for about 1 % of the total methane production during the growing season for grid cell A and B, and an even smaller percentage (average 0.6 % for grid cell A and B during summer) for the rhizospheric CH_4 oxidation. These leads to most of the methane that is produced to be emitted to the atmosphere through the different transport pathways. Past observational and laboratory studies have estimated the methane oxidation in boreal and tundra soils. Whalen and Reeburgh (2000) showed that about 55 % of the CH_4 diffusing from the saturated boreal soils, were oxidized while reaching the surface. Through bottle incubations, Knoblauch et al. (2016) measured the volumetric CH_4 oxidation potential of soil and moss samples collected from ponds of the Lena Delta. The fraction of produced CH_4 that is oxidized before it is emitted was then calculated following three different approaches. Their results show a mean fraction of produced CH_4 that was oxidized between 61 to 78 % estimated from a stable isotope approach, while slightly different values were found in samples from pond areas without vascular plants: up to 90 % of the CH_4 that was produced, was completely oxidized following a potential methanogenesis approach, and between 63 % to 94 % calculated from diffusive CH_4 fluxes into the bottom water.
Berestovskaya et al. (2005), measured CH_4 oxidation rates of different soil samples from the Russian Arctic tundra and found that generally the rates of methane oxidation exceeded those to the rates of methane production especially at temperatures of 5 degC. For this to happen, methane-oxidizing bacteria rapidly consumes the methane released from the freshly thawed tundra soils and the methane already deposited in the unfrozen soil, and this takes place even before methanogens produce new methane.
Based on these scarce observations in boreal soils, the oxidation processes in our model are still robust and need to be revisited in order to improve the contribution of the methane oxidation processes into the total methane emissions. We will discuss this in more detail in the revised ms.
It is important noting that the process-based model presented in this manuscript explicitly considers physical drivers such as soil moisture, inundated area, soil temperature and substrate availability for methane production and emissions, and is potentially one of the few models that include explicitly methane oxidation processes. However, despite our efforts of improving the process-based representation, intrinsic model shortcomings are still present, and those are related to setting a fixed criterion in the soil moisture to allow the accumulation of water above the surface, which leads to a loss in the connection to the soil temperature. This has been clearly stated in our manuscript. As such, a one-to-one comparison between the model results and observations can be hardly expected. Still, we demonstrated advances in the process-model based approach which lead to methane emissions results that are comparable in temporal variation in magnitude to those measured on site.

Line 61-62: Perhaps the authors could mention another important permafrost thaw effect on methane emissions here: increasing surface wetness due to surface subsidence of ice-rich soils (see for example Christensen et al., 2004; Johnston et al., 2014, Helbig et al., 2017).
OK, we will mention this process in the revised manuscript. Thanks for this suggestion.

Line 94-100: Wintertime methane emissions have also been reported by Helbig et al. (2017) for a boreal peat landscape in northwestern Canada, where they found winter emissions to contribute about 25 % to the annual budget.
Thank you for adding this citation that we overlooked because by the time our manuscript was submitted, this paper was still not published, but now it will be added.

Line 121: Could the authors discuss here the most important "shortcomings in the parameterization" of the state-of-the-art methane models?
The biggest limitations for modeling methane emissions in boreal regions are related to the complex network of processes with highly variable influences that are difficult to disentangle with temporally and spatially scarce field measurements. The available published literature is also scarce and focused only on fine scale site-level studies or at very coarse global scale. Several models that attempt to simulate methane emissions from soils are rather coarse and not well documented or evaluated, which leads the process-based principle remaining rather incomplete. These shortcomings are well documented in the cited paper of Xu et al., 2016. According to the methane models intercomparison exercise WETCHIMP (Bohn et al., 2015; cited in the discussion manuscript) the authors conclude that process-based methane models are limited by the following factors: availability of a valid and highly resolved wetland map, account for methane emissions and uptake in dry non-wetland areas, limited soil thermal physics that do not contain freeze and thaw processes, lack of a snow scheme and consequently, gas transport in the presence of snow cover, lack of peat soils.
From our work presented in this manuscript and our model development efforts in this topic, we have taken into account some of this shortcomings and improved our model tool, however still limitations exist. We still conclude that in boreal regions influenced by permafrost, process-based modeling for methane emissions is challenged by the lack of the observational measurements that can contribute e.g. to understand better the dynamics of soil moisture and temperature, wetlands distribution, as well as the distribution and temporal variation of roots in vascular plants. Additionally, the land surface models that serve as framework have intrinsic limitations in their design, e.g. in the case of JSBACH, the hydrology scheme does not allow the accumulation or horizontal redistribution of water and other tools such as TOPMODEL had to be implemented. Still after the inclusion of the TOPMODEL approach, the final methane emissions are restricted exclusively to the areas where standing water takes place, leaving out the dry areas to come into play, and TOPMODEL does not feedback the soil thermal physics. Finally, the carbon decomposition scheme in our JSBACH model version is only dependent on 15-day mean of air temperature and precipitation, which leads to an absence of permafrost carbon in this model version.
This piece of discussion will be completed in the revised ms to clarify better out statement of L121.

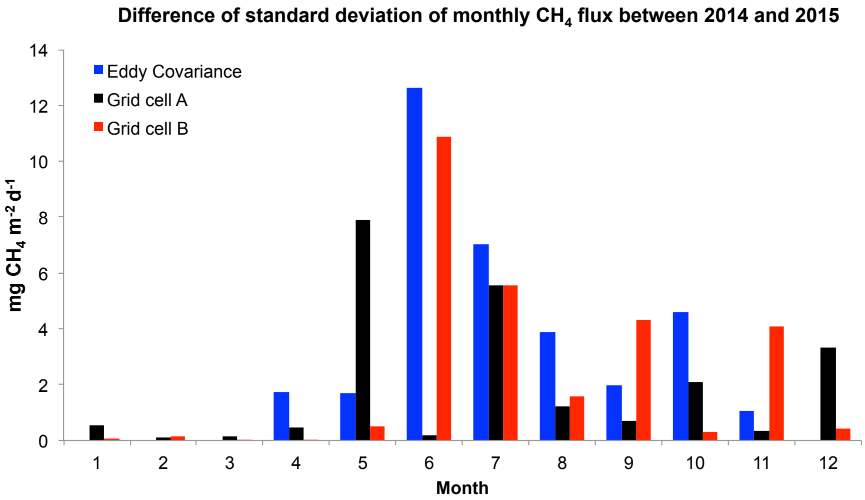
Line 133: Perhaps the work by Cresto-Aleina et al. (2013, 2016) on microtopography effects on surface water and methane emission dynamics could be mentioned here too.
OK, we will add this citation and process as suggested.

Line 500-501: Only mineral soils are considered for the methane modelling? How common are organic soil in the study area? I would assume that at least top-soils in the floodplain would be organic-rich. How would "considering" organic soils change the results?
The lacking representation of organic soils is a shortcoming JSBACH has in common with many other land surface models. The authors are only aware of two peatland-enabled versions of the LPJ (Lund-Potsdam-Jena) model in the published literature. In addition, a small number of further modelling studies have been published, where organic layers were

considered, though mainly for their thermal properties (e.g. Ekici et al., 2014, cited in discussion ms). This lacking representation is mainly due to the difficulties of coupling sub-gridscale hydrology and carbon cycle in a holistic way. The reviewer is right that organic soils are common in the study area. From measurements in the Chersky floodplain reported in Kwon et al., 2016, the soil layer has a top organic peat layer about 15-25 cm thick on top of alluvial material composed of silty clay. In the model configuration, only mineral soils are considered and indeed the organic carbon pools might be depressed in contrast to the organic carbon in a peat layer. This was discussed earlier above in this response.

Line 577-579: The authors may consider supporting this statement with information on the exact magnitude of interannual variability.

We calculated the magnitude of the interannual variability of the fluxes from eddy covariance and the model by comparing the standard deviation of the monthly values from 2014 to those in 2015. We summarize these results in the figure below.



The statement in the discussion ms is now supported by showing that largest interannual variability in the model grid cell A takes place in May and July with 7.9 and 5.5 mg CH_4 m^{-2} d^{-1} when compared the standard deviations from the monthly fluxes between 2014 and 2015, while for grid cell B, the largest variability between the two years took place in June and July (10.9 and 5.6 mg CH_4 m^{-2} d^{-1}, respectively). Still, the largest interannual variability was observed in June for the Eddy covariance data with 12.6 mg CH_4 m^{-2} d^{-1} difference in their monthly standard deviation between both years.

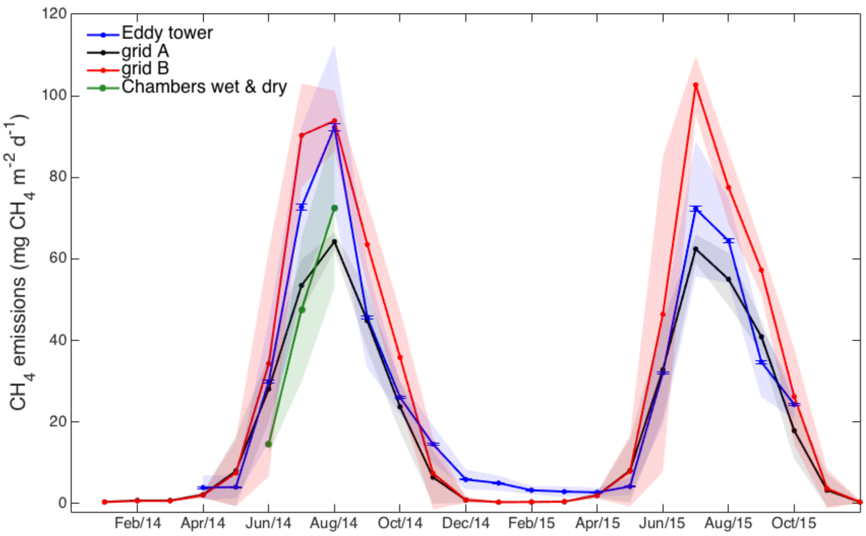
For practical reasons, we will complete the lines mentioned by the reviewer with the quantities obtained, and we will not show the figure as well.

Line 589-592: What is the uncertainty in the eddy covariance flux measurements? Could the authors quantify uncertainties due to random errors, gap-filling, u*-threshold, and footprint heterogeneity? An uncertainty quantification of eddy covariance fluxes would further strengthen the model-observation comparison.

The uncertainty analysis for the eddy-covariance flux data consists of random and systematic errors and is assessed based on well-established concepts (Aubinet et al., 2012).

Random errors linked to the turbulent sampling error and instrument error are given as standard output of the flux processing software TK3 (Mauder and Foken, 2015) for each 30 min flux value. Footprint uncertainties are not quantified, since there are no major transitions in biome types within the core areas of the flux footprints. Random errors are combined and considered as independent variables.

Systematic errors can occur due to unmet assumptions and methodological challenges, instrument calibration and data processing. Instruments are calibrated in regular intervals, and in comparison to a second eddy-covariance tower close by (~ 600 m) no systematic offset in the frequency distributions of wind speed, sonic temperature, and methane mixing ratios between towers was observed. The standardized software TK3 (Mauder and Foken, 2015) contains all the required processing steps for the flux data processing, as well as conversions and corrections, and yielded good agreement in a recent comparison with EddyPro (Fratini and Mauder, 2014). The post-processing quality control and flagging system scheme was based on stationarity and a well-developed turbulence scheme proposed by Foken and Wichura (1996) followed by additional tests applied to flag implausible data points in the resulting flux time series. Data coverage of methane fluxes was 86 % during the growing season and 67 % during the winter (Kittler et al., 2017).
The gap-filling method is based on a moving window that is centered in the gap and a 10-day window length, i.e. 5 days before and 5 days after the gap. The uncertainties were quantified as standard deviation for the corresponding window, similar to the gap-filling uncertainties for the CO_2 flux via the MDS routine (Reichstein et al., 2005).
No u*-threshold was applied to the flux dataset, since we determined the stationarity of the signal and integral turbulence characteristics also for nighttime conditions. This information facilitates identifying datasets with regular turbulent exchange also during stable stratification, therefore producing fewer gaps compared to a bulk exclusion of data during stable nighttime stratification through the u*-filter method. Random errors decrease with averaging and were calculated according to Rannik et al. (2016).
Averaged over both years (2014 and 2015) the CH_4 flux uncertainty based on 30 min data is 5.1±8.8 nmol m^{-2} s^{-1} (7±12.1 mg CH_4 m^{-2} d^{-1}). This result is not considering gap-filling techniques to the quality-checked signal (bulk uncertainty). The mean value considering also gap-filling is: 7.4±8.3 nmol m^{-2} s^{-1} (10.2±11.5 mg CH_4 m^{-2} d^{-1}).
For a fen ecosystem, it has been reported an uncertainty of 4.7±3.8 nmol m^{-2} s^{-1}. This result considers quality-checked data without applying a gap-filling technique (Jammet et al., 2017).
After considering monthly averaging of the gap-filling and with a quality checked signal, the uncertainties of the CH_4 fluxes measured from EC for 2014 and 2015 are reduced to 0.35±0.22 mg CH_4 m^{-2} d^{-1}. Monthly uncertainty values will be included in Figure 5 for the revised ms as error bars of the mean monthly values. The updated figure is presented below. Details on data uncertainty assessment as outlined above, will be provided in a new Appendix B on "Details on in-situ flux observations". References cited in this section are listed at the end of this response.



Line 691-711: I am not sure how this section contributes to the research questions of this manuscript? Perhaps the authors could mention differences in environmental characteristics of grid-cell A and B briefly in the manuscript and move figure 9 to the supplementary material?
We will shorten this section and instead merge it with the discussion of methane fluxes, in this way we could move figure 9 to the supplementary material. This suggestion certainly will make the manuscript more focused on the main aim. Thanks for this suggestion.

Line 808-810: The impact of cooler early summer temperatures on soil warming and methane emissions has been demonstrated recently using multi-year methane observations in a boreal peat landscape (see Helbig et al., in press). The authors may consider discussing their modelling results in relation to these observations.
We thank the reviewer for making us aware of this new publication. We will add it in the revised ms and cite it accordingly. Helbig et al., 2017 shows that between years 2013 and 2016, during May of each year the one in 2014 was colder compared to the other years. This finding was based on a meteorological record of an area in northwestern Canada. As a result of temperature shifts, soil temperatures varied and influenced the year-to-year methane fluxes, specially variations in spring soil temperature were influential. The findings of Helbig et al. are in good agreement with our model observations regarding the interannual variability in air and soil temperature and their influence in methane emissions. We will complete our model results with this nice comparison.

Line 847-851: The authors may consider starting the discussion mentioning the parameters that actually made a difference and not with the parameters that did not change the results. It should be highlighted what process/parameter matters in the model.
Thank you for this suggestion. We will re-structure the discussion based on this suggestion.

Line 991-992: Few studies have shown that non-inundated upland areas may take up methane (e.g., Flessa et al., 2008). As far as I understand, such uptake is not considered in the current work. How could uptake in the drier areas of the model domain change simulation results? There are large areas in the model domain that appear to be characterized by upland landscapes and thus potential methane uptake (see Fig. 1).
Indeed atmospheric CH_4 uptake should not be neglected when considering a regional CH_4 budget, especially when the majority of areas are predominantly aerobic. With plot-based observations in dry areas of the Chersky floodplain, the CH_4 emissions where negative, indicating uptake (average of -3 mg CH_4 m^{-2} d^{-1} during summer of 2014) and this value was considerably smaller compared to that of the CH_4 emissions measured in wet plots (on average 332 mg CH_4 m^{-2} d^{-1} in summer 2014) (see response above for figure of results). Based on this result and the consideration of an inundated fraction of 20 % during summer in the model grid cell A (Fig. 9d in discussion manuscript), about 66.4 mg CH_4 m^{-2} d^{-1} (in wet plots: 332 mg CH_4 m^{-2} d^{-1} * 0.2 = 66.4 mg CH_4 m^{-2} d^{-1}) are emitted and 2.4 mg CH_4 m^{-2} d^{-1} are loss by uptake (oxidized) (given by the dry plots results: -3 mg CH_4 m^{-2} d^{-1} * 0.8 = -2.4 mg CH_4 m^{-2} d^{-1} equivalent to 8.7 % of the total methane emissions), leading to the net CH_4 emission of 64 mg CH_4 m^{-2} d^{-1} according to chamber measurements. The mean CH_4 emission in grid cell A during June-July-August 2014 (Fig. 5a of discussion manuscript) is 48.6 mg CH_4 m^{-2} d^{-1} in the inundated areas. The mean methane soil and plant oxidation for the same period of time, given by the model for grid cell A, is 0.63 mg CH_4 m^{-2} d^{-1} (0.3 % of the total emission) which is low compared to the uptake estimation for the chamber measurements in dry areas. However, these are only representing oxidation processes in saturated soils, which are not predominant in contrast to dry soils. As mentioned before, by not considering non-inundated areas in the modeling of methane processes, the methane uptake is ultimately underestimated because the conditions for methane oxidation are limited. The model can be further improved in the CH_4 oxidation scheme, but this can only be possible after a thorough observation of CH_4 uptake rates and their controlling factors in this area, and also as the

hydrology scheme is also improved. More on the methane oxidation in the model is discussed in this response and will be also emphasized in the revised ms.

Line 1134-1141: The authors may consider not to introduce a new concept (e.g., anaerobic microsites) at the very end of the conclusions. I would recommend to only refer here to what has been shown in the manuscript so far.
OK, we will improve this section in the revised manuscript.

Line 1252-1255: What would happen if the model would run with the old order of processes? Shouldn't this be part of the uncertainty analysis?
The old order of processes was presented in the paper by Kaiser et al., 2017; there, it was shown that the order of processes was selected based on the velocity that they physically can exhibit, with ebullition first and the slowest transport at last which was plant mediated transport due to the resistance of the plants exodermis. Observational evidences however, as discussed here and in the manuscript, show that in the presence of vascular plants, wetland annual methane emissions are mainly from the transport of gas through plants. Due to the structure of the model, it is not possible to run parallel processes and instead, a sequential flow of processes has to be computed. For this reason, the solution to improve the individual share of the transport processes was to re-arrange the processes by expected priority. We do not think this should be part of the sensitivity studies for this manuscript since this is a purely computational design and not due to the inherent processes in the model.

Fig. 1: Why did the authors use such a large study area, if ground-based observations were only available for a very small fraction of the model domain? How can the model performance be evaluated for the other non-floodplain grid cells that appear to be characterized by different landscape characteristics?
We agree with the reviewer that a smaller study area could have been shown, especially for the area near the grid cell where ground-based observations take place.
Plot level model simulation have been performed in the past, particularly with a similar version of the model presented in this manuscript for a site in Samoylov (Kaiser et al., 2017). Model development for methane emissions has not only focused on the improvement in the mechanisms represented in the model for the production and transport of methane, but also in the scaling with the intention of understanding better the contribution of CH_4 processes over larger spatial scales. Regional scales still pose a challenge but certainly models need to be aimed to be applied to larger scares rather than only plot level. After a plot level application, we improved the description of some processes in the model and aim to test it in a rather larger, but still, regional spatial scale.
Thus, the intention of selecting a larger regional domain is two-fold: 1) to test and apply the process-based methane model in a larger than site-level domain and 2) to identify the heterogeneity in the methane processes linked to different soil and vegetation conditions, this is important since sub-grid soil heterogeneity is still not represented in the model, and is also particularly relevant for large-scale inundation evaluation.
We agree with the reviewer that no observational data is available to evaluate more than one model grid cell, and indeed one should be careful with the interpretation of the non-floodplain areas of the model domain, however, our contribution here is also aimed to be used for even larger domains and for future predictions, so testing such model already in larger scale and been showing its computational capability and overall realistic performance is a step forward towards that aim. As more observational efforts will be done in the future, in other areas near the Kolyma region and Chersky floodplain for our own practical purposes, the model will be able to be evaluated for those other areas. Also, an intercomparison between JSBACH results and atmospheric inverse modeling at the regional scale is in preparation. Until then, we still believe in the value of our scientific contribution to evidence the applicability of a refined process-based methane model.

Fig. 6: Why do the authors compare the mean grid-cell soil temperature profile to measured wet and dry soil temperature profiles? Physical soil properties differ drastically between wet and dry soils and consequently strongly determine soil temperature dynamics (see end of discussion). Wouldn't it be therefore necessary to at least model soil temperature dynamics of the inundated and non-inundated land surface separately?

We present these data in Fig. 6 to show the existing model physical state that was used for the calculation of methane emissions in the model. We agree with the reviewer that ideally, the model should be able to produce results separately for the dry and for the wet soil areas. However and unfortunately, this is not possible with the current model configuration and this is due to the basic model structure of JSBACH. Each model grid cell is subdivided into tiles that only serve to describe different vegetation types, however the soil properties remain the same for the entire grid cell and average soil state variables are considered. Thus, the soil temperature dynamics actually represent the entire grid cell and these are independent of the TOPMODEL interactions, i.e. inundated and non-inundated areas. This is obviously a shortcoming in this version which was presented in the discussion ms and which is true for many other land surface models. In order to represent sub-grid heterogeneity of soil properties the model configuration would need to be completely restructured which we hope can be done in the near future with the new developments of the JSABCH 4.0.

Fig. 7: Methane emissions increase considerably in the model at sub-zero soil temperatures. In contrast, measured methane emissions appear to be quite insensitive to soil temperature below 0_C. The authors mention this mismatch in lines 655-659. Perhaps the authors can discuss this mismatch between temperature-emission responses in more detail. How is it possible that such cold simulated soil temperatures result in emissions of > 30 mg CH4 m-2 day-1?

We agree with the reviewer that wintertime processes are still not well captured with our current model configuration. This goes down basically to the soil moisture that had to be artificially modified to allow the accumulation of water at the soil surface according to the topographic profile. The results presented in Fig. 7 show that, the high model methane emissions mentioned by the reviewer, take place mostly during October and May (grey circles and triangles for grid cells A and B, respectively) and this reflect the gradual transition of the emissions as the soil starts to freeze towards December and also as it starts to melt before summer. Comparing with the observations, this result seems implausible, however, we think is not also impossible to happen. In the work of Zona et al. (2016), the authors demonstrated the emissions of methane during the zero curtain period. In their Figure 3, panel B, high methane emissions take place still at subzero temperatures (on average 7.8 mg CH_4 m^{-2} d^{-1} at -5 degC) between September and December in 2014, while in panel A, the methane fluxes behave more similarly to our observations in the Chersky floodplain (barely changing < 0 degC). Still the magnitude of the observed emissions is not as large as what we observe with JSBACH and here the model parameters and schemes might play the role. The zero curtain period presented in Zona et al. reflects the release of CH4 still in autumn, due to the production of CH4 in sub-soil warm layers. To investigate if the results of our model reflect somehow this process as well, still other schemes in the model must be revisited and improved such as: Q10 and water impact in carbon decomposition, and processes such as soil freezing under moisture limitation and thermal soil response.

Fig. 8: Here, an uncertainty estimate for the measured cumulative methane emissions would help interpreting the comparison between simulated and measured fluxes.

In order to include uncertainty estimates to the cumulative methane emissions presented in Fig. 8, we calculated the monthly cumulative fluxes in panel e and added the error bars as standard deviation of the monthly cumulative fluxes. Despite our discussion regarding the total cumulative fluxes when comparing the eddy covariance record to the model grid cells results, we observe that the uncertainty in the monthly fluxes is larger in all of the data sets during October 2014 and generally decreases toward April 2015. The uncertainty ranges are

also generally larger in the eddy covariance data and this is due to the high intrinsic signal daily variability.
We updated this figure in the revised ms and discussed the uncertainty values. The new figure is:

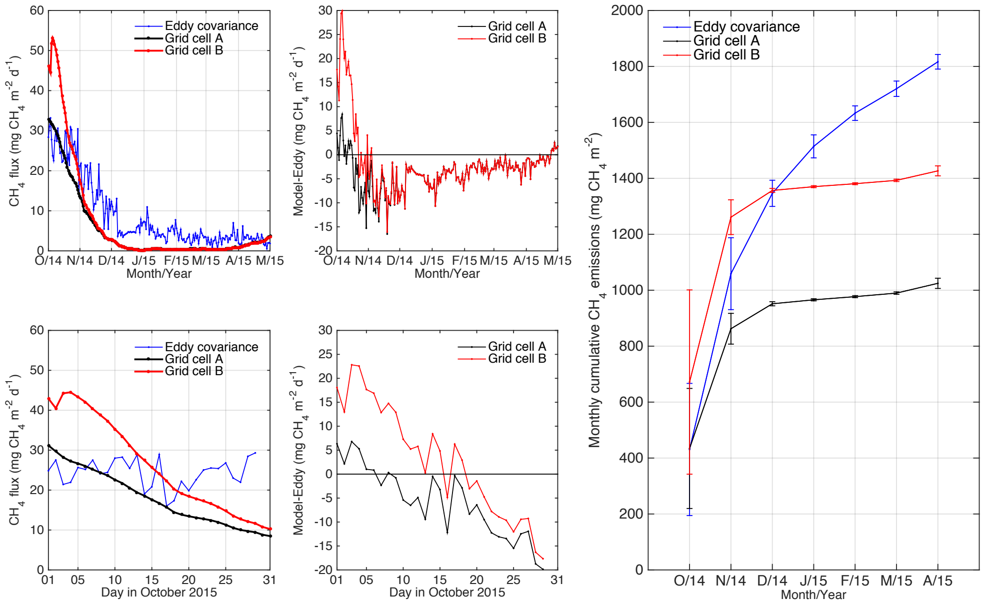


Fig. 11: I am not sure how this figure contributes to the research questions. The seasonality of different methane emission pathways is already shown in Fig. 10. How does a representation of the spatial distribution of the methane emissions add to the manuscript?
We still think that showing the total model domain methane emissions from the different pathways evidence the skill of the model for its regional application. In Fig. 10, only grid cells A and B are shown. Furthermore, by following the suggestion of the reviewer of moving Fig. 10 to supplementary material, we still would like to keep Fig. 11 as part of the main ms.

Technical comments

Line 149: Remove “done”.
OK
Line 150: Remove “are”.
OK
Line 196: Please define what “hospitable and inhospitable” land means in this context.
We have completed this paragraph by adding the following lines: “A prescribed fraction of each grid cell is used to discriminate between land hospitable and inhospitable to vegetation. In JSBACH, each grid cell has a designated fraction where vegetation cover types across tiles can be assigned, hence is the fraction hospitable to vegetation. The remaining fraction of the grid cell is then associated to a land cover type that represents areas where vegetation does not grow, such as rocky surfaces and deserts; hence it is considered inhospitable to vegetation (Reick et al., 2013).

Line 534: What do the authors mean with “visually”? They state in the previous sentence that differences are not statistically significant.
We refer here to the time shift in the mean methane emissions signal when the sensitivity experiments are compared. However, indeed the statistical analysis showed that there is no significant difference between the results of the sensitivity tests for the individual and total emissions. We rephrased this sentence to avoid confusion, it now reads: “A time shift is seen however, in the CH_4 emissions from mid-October until mid-November (Fig. 3, column 4 of row e), with the larger emissions through snow taking place earlier if h_{snow} is thinner.

Nevertheless, this temporal shift in the CH_4 emissions through the snow is not observed in the total CH_4 emissions."

Fig. 3: Please clarify what the inset figures show.
Thank you for pointing this out. Now we added the following sentence to the caption: "The inset figures in some of the panels are zooms to periods of time where larger difference between signals is depicted."

References cited in this response that are not included in the original discussion manuscript:

Aubinet, M., T. Vesala, and D. Papale (2012), Eddy Covariance - A practical guide to measurement and data analysis, 449 pp., Springer, Dordrecht; Heidelberg; London; New York, doi:10.1029/2002JD002055.

Berestovsakaya, Y.Y., I.I. rusanov, l.V. Vasil'eva and N.V. Pimenov (2005), The processes of methane production and oxidation in the soils of the Russian Arctic tundra, Microbiology, 74(2), 221-229, 10.1007/s11021-005-0055-2.

Chapin III, F.S. and M. Slack (1979), Effect of defoliation upon root growth, phosphate absorption and respiration in nutrient-limited tundra graminoids, Oecologia, 42, 1, 67-79, 10.1007/BF00347619.

Foken, T., and B. Wichura (1996), Tools for quality assessment of surface-based flux measurements, Agr Forest Meteorol, 78(1-2), 83-105, doi:10.1016/0168-1923(95)02248-1.

Fratini, G., and M. Mauder (2014), Towards a consistent eddy-covariance processing: an intercomparison of EddyPro and TK3, Atmos Meas Tech, 7(7), 2273-2281, doi:10.5194/amt-7-2273-2014.

Jammet, M., Crill, P., Dengel, S., and Friborg, T.: Large methane emissions from a subarctic lake during spring thaw: mechanisms and landscape significance, Journal of Geophysical Research: Biogeosciences, 120, 2289-2305, 2015, 10.1002/2015JG003137.

Jammet, M., S. Dengel, E. Kettner, F. J. W. Parmentier, M. Wik, P. Crill, and T. Friborg (2017), Year-round CH_4 and CO_2 flux dynamics in two contrasting freshwater ecosystems of the subarctic, Biogeosciences, 2017, 14, 5189-5216, doi:10.5194/bg-14-5189-2017.

Kittler, F., M. Heimann, O. Kolle, N. Zimov, S. Zimov, and M. Göckede (2017), Long-term drainage reduces CO_2 uptake and CH_4 emissions in a Siberian permafrost ecosystem, Global Biogeochem. Cy., 31, 1704-1717, doi:10.1002/2017GB005774.

Mauder, M., and T. Foken (2015), Eddy-Covariance Software TK3, doi:10.5281/zenodo.20349.

Rannik, U., O. Peltola, and I. Mammarella (2016), Random uncertainties of flux measurements by the eddy covariance technique, Atmos Meas Tech, 9(10), 5163-5181, doi:10.5194/amt-9-5163-2016.

Reichstein, M., et al. (2005), On the separation of net ecosystem exchange into assimilation and ecosystem respiration: review and improved algorithm, Glob Change Biol, 11(9), 1424-1439, doi:10.1111/j.1365-2486.2005.001002.x.

Reick, C., T. Raddatz, V. Brovkin, V. Gayler, (2013), Representation of natural and anthropogenic land cover change in MPI-ESM, Journal of Advances in Modeling Earth Systems, 5, 459-482, 10.1002/jame.20022.

Saltelli, A., Chan, K. & Scott, E. M. (2000), *Sensitivity Analysis. Wiley Series in Probability and Statistics*, John Wiley & Sons, ltd.

Tokida, T., Mizoguchi, M., Miyazaki, T., Kagemoto, A., Nagata, O., and Hatano, R.: Episodic release of methane bubbles from peatland during spring thaw, Chemosphere, 70, 165-171, 2007, 10.1016/j.chemosphere.2007.06.042.

Wang, P., L. Mommer, J. van Ruijven, F. Berendse, T. C. Maximov, M. M. P. D. Heijmans, 2016, Seasonal changes and vertical distribution of root standing biomass of graminoids and shrubs at a Siberian tundra site, Plan Soil, 10.1007/s11104-016-2858-5.

Whalen, S.C. and W.S. Reeburgh (2000), Methane oxidation, production, and emission at contrasting sites in a Boreal Bog. Geomicrobiology Journal, 17, 237-251.

References recommended by reviewer (and added also to the revised ms).

Christensen TR, Johansson T, Åkerman HJ et al. (2004) Thawing sub-arctic permafrost: Effects on vegetation and methane emissions. Geophysical Research Letters, 31, L04501.

Cresto-Aleina F, Brovkin V, Muster S, Boike J, Kutzbach L, Sachs T, Zuyev S (2013) A stochastic model for the polygonal tundra based on Poisson – Voronoi diagrams. Earth System Dynamics, 4, 187–198.

Cresto-Aleina F, Runkle BRK, Brücher T, Kleinen T, Brovkin V (2016) Upscaling methane emission hotspots in boreal peatlands. Geoscientific Model Development, 9, 915–926.

Flessa H, Rodionov A, Guggenberger G et al. (2008) Landscape controls of CH4 fluxes in a catchment of the forest tundra ecotone in northern Siberia. Global Change Biology, 14, 2040–2056.

Helbig M, Boike J, Langer M, Schreiber P, Runkle BRK, Kutzbach L (2013) Spatial and seasonal variability of polygonal tundra water balance: Lena River Delta, northern Siberia (Russia). Hydrogeology Journal, 21, 133–147.

Helbig M, Chasmer L, Kljun N, Quinton WL, Treat CC, Sonnentag O (2017) The positive net radiative greenhouse gas forcing of increasing methane emissions from a thawing boreal forest-wetland landscape. Global Change Biology, 23, 2413–2427.

Helbig M, Quinton WL, Sonnentag O (2017) Warmer spring conditions increase annual methane emissions from a boreal peat landscape with sporadic permafrost. Environmental Research Letters, 12, 115009, doi: 10.1088/1748-9326/aa8c85.

Johnston CE, Ewing S a, Harden JW et al. (2014) Effect of permafrost thaw on CO2 and CH4 exchange in a western Alaska peatland chronosequence. Environmental Research Letters, 9, 85004.

Liljedahl AK, Boike J, Daanen RP et al. (2016) Pan-Arctic ice-wedge degradation in warming permafrost and influence on tundra hydrology. Nature Geoscience, 9, 312–318.

Parmentier FJW, van Huissteden J, van der Molen MK, Schaepman-Strub G, Karsanaev SA, Maximov TC, Dolman AJ (2011) Spatial and temporal dynamics in eddy covariance observations of methane fluxes at a tundra site in northeastern Siberia. Journal of Geophysical Research, 116, G03016.

Sachs T, Giebels M, Boike J, Kutzbach L (2010) Environmental controls on CH4 emission from polygonal tundra on the microsite scale in the Lena river delta, Siberia. Global Change Biology, 16, 3096–3110.

Appendix B: Details on in-situ flux observation program

Eddy-covariance flux data uncertainty assessment

Following well-established procedures in literature (Aubinet et al., 2012), our uncertainty analysis for the eddy-covariance flux data has been split up into random and systematic errors. The major sources for random errors, associated with the turbulent sampling and instrument issues, have been quantified for each 30 min flux value through the flux processing software TK3 (Mauder and Foken, 2015). Errors related to footprint uncertainties were not quantified, since there are no major transitions in biome types within the core areas of the flux footprints.

Systematic errors can be introduced by unmet theoretical assumptions and methodological challenges, as well as by instrument calibration and data processing issues. In the context of the Chersky observations, instruments were maintained and calibrated in regular intervals, therefore minimizing this potential error component. Moreover, data intercomparisons with a second eddy-covariance tower close by (~ 600 m) yielded no systematic offset in the frequency distributions of wind speed, sonic temperature, and methane mixing ratios between towers. Regarding flux data processing, the TK3 software package contains all the required processing steps, conversions and corrections for the flux data processing, and yielded good agreement in a recent comparison with EddyPro (Fratini and Mauder, 2014). To avoid methodological issues that may bias flux data results, we employed a rigid post-processing quality control and flagging system scheme, with the well-established analyses for stationarity and well-developed turbulence originally proposed by Foken and Wichura (1996) at its core, supplemented by additional tests (absolute range and spikes) to flag implausible data points in the resulting flux time series. Based on the quality assessment and control tools outlined above, we excluded systematic errors from the uncertainty quantification of flux data that were assigned a high to medium data quality (QF 1-6 based on the scheme proposed by Foken et al., 2005; 2012) and subsequently used for assessing long-term CH_4 flux budgets.

No u*-threshold was applied to the flux dataset, since we determined stationarity of the signal and integral turbulence characteristics also for nighttime conditions. This information facilitates identifying datasets with regular turbulent exchange also during stable stratification, therefore producing fewer gaps compared to a bulk exclusion of data during stable nighttime stratification through the u*-filter method. After filtering out low-quality fluxes, the data coverage of methane fluxes was 86 % during the growing season and 67 % during the winter from the original full 30 min flux data set (Kittler et al., 2017). To produce a continuous flux record for quantification of long-term CH_4 budgets, we filled the remaining gaps by averaging existing flux data within a moving window of 10-day length centered on the gap. Uncertainties for gap-filled values were quantified as standard deviation within the corresponding window, similar to the definition of gapfilling uncertainties for the CO_2 flux via the well-established marginal distribution sampling routine by Reichstein et al. (2005).

To produce aggregated uncertainty values for longer time periods, we applied the procedures suggested by Rannik et al. (2016). All random errors were combined by considering them as independent variables, and normally decrease with the length of the averaging period. Averaged over both data years used within the context of this study (2014 and 2015), the CH_4 flux uncertainty based on the 30 min data is 7.4±8.3 nmol m^{-2} s^{-1}, a result comparable to 4.7±3.8 nmol m^{-2} s^{-1} reported for a fen ecosystem by Jammet et al. (2017).

Source weight function of the eddy-covariance flux data

We conducted a source weight analysis, also called footprint analysis, to determine the fractional contribution of different land cover types within the field of view of the eddy-covariance flux tower. Source weight functions for each 30min flux measurement were computed based on the Lagrangian Stochastic footprint model by Rannik et al. (2003). Footprints were accumulated, analyzed and interpreted using an approach presented by Göckede et al. (2006; 2008). We projected these footprints onto a WorldView2 land cover map at 2 m horizontal resolution (see also Figure A1). In the context of the presented study, we aggregated the original 22 land cover classes into 9 classes to concentrate on the dominant elements of the vegetation community structure (see also Table A1).

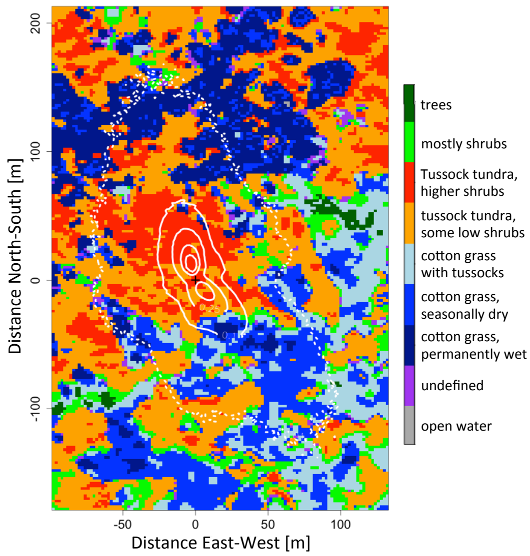


Figure A1: Accumulated source weight function for the control tower within the Chersky study site, based on data from the growing season (mid June – mid September) in 2014. Solid white isolines indicate the 80, 60, 40, and 20% levels, while the dashed line gives the 10% level. Background colors give aggregated land cover classes based on WorldView2 data.

Since the tower is situated on a slightly elevated patch of tundra, tussocks and shrubs featuring various levels of wetness (red and orange colors in Fig. A1) dominate the immediate surroundings. Even though inundated parts of the study area, in this case identified by the prevalence of the cotton grass *Eriophorum angustifolium* (blue-ish colors in Fig. A1), are dominating the area encircled by the 10% isoline that is used here to mark the boundary of the cumulative footprint area, they are mostly present in the outer reaches, therefore combining just about 26% of the total flux signal sampled by the eddy system. Another 31% is contributed by wet to moist tussock tundra with some shrubs. Overall coverage fractions within the major wetness categories (see also Table A1) remain approximately constant between tower footprint and two larger regions covered by the same WorldView dataset, indicating that this composition of wetness levels is typical for the Kolyma floodplain ecosystems analyzed within the context of this study.

Table A1: Fractional coverage of aggregated WorldView land cover classes within the control tower footprint of the Chersky study site. Background color coding was used to categorize the classes into wetness levels. The rightmost two columns give fractional coverage of these classes within the area immediately surrounding the towers (1.2 x 1.2 km) and within the entire WorldView scene analyzed (5 x 5 km).

Land cover class	category	tower footprint	1.2x1.2km	5x5km
water	open water	0.001	0.035	0.134
cotton grass, wet continuously		0.111	0.043	0.053
cotton grass, partially dry	wetland	0.067	0.153	0.147
cotton grass with tussocks		0.081	0.063	0.038
tussocks with some shrubs	wet to moist	0.312	0.418	0.280
tussocks with higher shrubs		0.388	0.165	0.182
higher shrubs, with tussocks	moist to dry	0.031	0.097	0.115
trees		0.001	0.006	0.017
undefined		0.008	0.020	0.035

Flux chamber observations

The Chersky study site features two transects of 10 permanently installed PVC collars for flux chamber measurements. With distances of approximately 25 m between individual microsites, both transects cover a distance of ~225 m within the drained and control sections, of this permafrost site. Site locations were selected quasi-randomly to reflect the dominant microsite characteristics (e.g. vegetation composition, wetness level) found at each of the target locations. With a chamber footprint of 60 cm x 60 cm, this technique allowed studying microsites with rather homogeneous environmental conditions, as compared to the eddy-covariance fluxes with often heterogeneous footprint areas. Details on the chamber program, overall methane flux rates observed, and functional relationships with e.g. soil temperature, vegetation and wetness levels, are provided by Kwon et al. (2016; 2017).

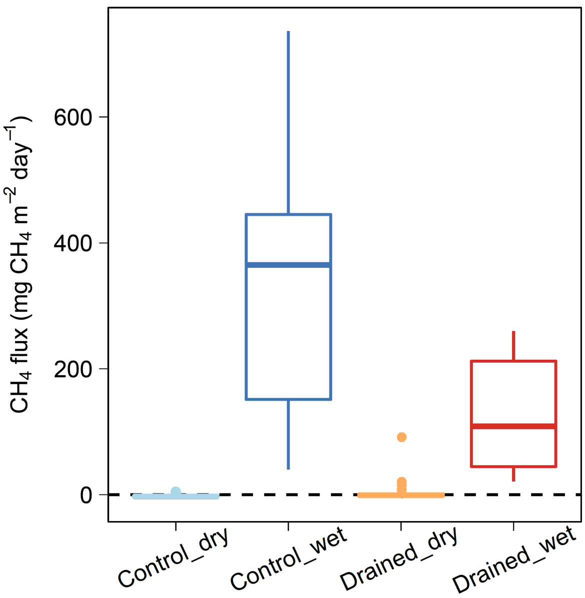


Figure A2: Daily methane flux rates aggregated from flux chamber measurements within the growing season of 2014. Measurements are separated into drained (1 wet microsite, 9 dry microsites) and control (8 wet microsites, 2 dry microsites) transects.

Figure A2 displays average flux rates for wet and dry microsites observed within the drained and control transects during sampling campaigns in summer 2014. These flux chamber results clearly demonstrate that methane release rates were virtually zero in the absence of standing water. At some of the dry microsites (results not shown), even negative CH_4 flux rates were observed, indicating the oxidation of methane under highly aerobic conditions within these predominantly wet tussock tundra ecosystems in Northeastern Siberia.

Cited literature

Aubinet, M., T. Vesala, and D. Papale (Eds.) (2012), Eddy Covariance - A practical guide to measurement and data analysis, 438 pp., Springer, The Netherlands.

Foken, T., R. Leuning, S. P. Oncley, M. Mauder, and M. Aubinet (2012), Corrections and data quality, in Eddy Covariance - A practical guide to measurement and data analysis, edited by M. Aubinet, T. Vesala and D. Papale, pp. 85-131, Springer, Dordrecht; Heidelberg; London; New York.

Foken, T., M. Göckede, M. Mauder, L. Mahrt, B. Amiro, and W. Munger (2005), Post-Field Data Quality Control, in Handbook of Micrometeorology: A Guide for Surface Flux Measurement and Analysis, edited by X. Lee, W. Massman and B. Law, pp. 181-208, Springer Netherlands, Dordrecht, doi:10.1007/1-4020-2265-4_9.

Foken, T., and B. Wichura (1996), Tools for quality assessment of surface-based flux measurements, Agr. Forest Meteorol., 78(1-2), 83-105, doi:10.1016/0168-1923(95)02248-1.

Fratini, G., and M. Mauder (2014), Towards a consistent eddy-covariance processing: an intercomparison of EddyPro and TK3, Atmos Meas Tech, 7(7), 2273-2281, doi:10.5194/amt-7-2273-2014.

Göckede, M., et al. (2008), Quality control of CarboEurope flux data - Part 1: Coupling footprint analyses with flux data quality assessment to evaluate sites in forest ecosystems, Biogeosciences, 5(2), 433-450.

Göckede, M., T. Markkanen, C. B. Hasager, and T. Foken (2006), Update of a footprint-based approach for the characterisation of complex measurement sites, Bound.-Lay. Meteorol., 118(3), 635-655, doi:10.1007/s10546-005-6435-3.

Jammet, M., S. Dengel, E. Kettner, F. J. W. Parmentier, M. Wik, P. Crill, and T. Friborg (2017), Year-round CH4 and CO2 flux dynamics in two contrasting freshwater ecosystems of the subarctic, Biogeosciences, 14(22), 5189-5216, doi:10.5194/bg-14-5189-2017.

Kittler, F., M. Heimann, O. Kolle, N. Zimov, S. Zimov, and M. Göckede (2017), Long-term drainage reduces CO2 uptake and CH4 emissions in a Siberian permafrost ecosystem, Glob. Biogeochem. Cy., (online first), doi:10.1002/2017GB005774.

Kwon, M. J., M. Heimann, O. Kolle, K. A. Luus, E. A. G. Schuur, N. Zimov, S. A. Zimov, and M. Göckede (2016), Long-term drainage reduces CO2 uptake and increases CO2 emission on a Siberian floodplain due to shifts in vegetation community and soil thermal characteristics, Biogeosciences, 13(14), 4219-4235, doi:10.5194/bg-13-4219-2016.

Kwon, M. J., et al. (2017), Drainage enhances surface soil decomposition but stabilizes old carbon pools in tundra ecosystems Nat. Clim. Change, (submitted).

Mauder, M., and T. Foken (2015), Documentation and Instruction Manual of the Eddy-Covariance Software Package TK3Rep., University of Bayreuth, Bayreuth.

Rannik, U., T. Markkanen, J. Raittila, P. Hari, and T. Vesala (2003), Turbulence statistics inside and over forest: Influence on footprint prediction, Bound.-Lay. Meteorol., 109(2), 163-189, doi:10.1023/a:1025404923169.

Rannik, Ü., O. Peltola, and I. Mammarella (2016), Random uncertainties of flux measurements by the eddy covariance technique, Atmos. Meas. Tech., 9(10), 5163-5181, doi:10.5194/amt-9-5163-2016.

Reichstein, M., et al. (2005), On the separation of net ecosystem exchange into assimilation and ecosystem respiration: review and improved algorithm, Glob Change Biol, 11(9), 1424-1439, doi:10.1111/j.1365-2486.2005.001002.x.