Dear Editor,

Thank you for the chance to respond to the remaining minor comments. We have addressed the reviewer comments and edited the manuscript where changes were requested. Please find our response to the reviewer below. We have copied the reviewer comments and our responses are in blue, with revised test and line numbers further indicated in italics below our responses as applicable.

Thank you,
Sarah Ludwig (on behalf of all authors)

Reviewer comments:
This is a review of the revised manuscript version. The authors have provided response to my original review comments and revised the manuscript accordingly. Overall, I find the response satisfactory and am pleased to recommend publication. I have some further, relatively minor comments the authors may wish to consider (RE-REVIEW COMMENTs below). I do not expect an additional review round.

(1)

REVIEW COMMENT - lines 217-218: Again, the authors have an idea how methane fluxes behave but do not reveal the source of this information. The Bayesian prior selection should be justified more carefully. Disallowing methane uptake clearly affects the posterior methane flux distribution of the degraded permafrost (Fig. 5).

AUTHORS' RESPONSE - The more specific prior information used here came from the archived dataset of chamber-based methane fluxes from these landcovers, which is cited here.

RE-REVIEW COMMENT:

The original comment referred to the following sentence: "We used mostly uninformative prior fluxes for landcovers anticipated to support CH4 emissions by disallowing CH4 uptake for degraded, edge of degraded, wetland, and water landcover classes". The dataset mentioned in the authors' response (Ludwig et al., 2018) does not follow this sentence but the following one, which is about the peat plateau fluxes. The dataset in question contains CH4 flux data only for the peat plateau (divided into lichen and moss patches, 36 data points in total). As far as I can see, there are no flux data for the wetland and water landcover classes. Furthermore, it is not obvious which data included in the dataset correspond to the 'degraded' and 'edge of degraded' classes as the classification is not fully consistent with the manuscript. It would be useful to present a consistent, quantitative summary of the measurement data that were actually used in the Bayesian prior selection process.

Thank you for clarifying. We have added a citation for fluxes from another dataset from water, wetland, and degraded permafrost sites nearby that clearly demonstrate these are methane producing locations. The assumption that wetlands and small ponds do not take up methane is very reasonable and does not require specific quantification to be summarized here. We have further clarified in the text that these sites are fully saturated soils or open water.

Lines 215-218: We used mostly uninformative prior fluxes for landcovers known to emit CH_4 (e.g. fully saturated soils and open water) by disallowing CH_4 uptake for degraded, edge of degraded, wetland, and water landcover classes (Ludwig et al. 2018a).

(2)

REVIEW COMMENT - lines 244-245: Here, the flux independency of biometeorological drivers is presented as if it were shown above in the manuscript ("Since the CH4 fluxes did not have relationships…").

AUTHORS' RESPONSE - We have added a figure to the SI (Fig S2 in the response above) demonstrating the lack of relationship between methane fluxes and biometeorological variables.

RE-REVIEW COMMENT:

The new figure (Fig. S2) aims to demonstrate that the CH4 flux does not depend on environmental drivers. Perhaps the authors could explain why they have chosen PAR and air temperature as the drivers that would potentially control the CH4 flux (rather than soil temperature and moisture) and why they aggregate the data for the April-September period while the fluxes are modelled on a monthly basis. In any case, the statement that "CH4 fluxes did not have relationships to any biometeorological drivers such as air temperature or PAR (Fig S2)" (line 253) should be revised as the relationship was only tested for these two variables.

Soil temperature and moisture were not measured in wetlands, degraded permafrost, or edge of degraded permafrost during this study period. We have change 'any' to 'observed' in the main text. The figure S2 does not show aggregated data, it shows scatterplots of all observations. Whether within months or across the growing season, there is no relationship between CH4 and PAR or temperature.

Lines 243: Since the CH4 fluxes did not have relationships to observed biometeorological drivers such as air temperature or PAR (Fig. S2)…

(3)

REVIEW COMMENT - What is the rationale behind calculating the carbon budget as CO2-equivalents where methane fluxes are multiplied by 28 (lines 265-266), which refers to the global warming potential due to a pulse emission over a time horizon of 100 years. What does this quantity indicate in this context? Why a 100-year period? Why a pulse emission approach for a natural ecosystem with fluxes sustained for thousands of years? Also, it is somewhat misleading to call the resulting sum "carbon budget" (that would logically be CO2-C + CH4-C).

AUTHORS' RESPONSE - We have chosen to do this because these systems are highly subject to climate warming and there is wide interest in interpreting methane emissions on the same scale of impact as carbon dioxide (as seen in Bastviken et al., 2011; Stocker, 2013; Euskirchen et al., 2014; Beaulieu et al., 2020; Skytt et al., 2020). While the 28 global warming potential is not a perfect representation of methane emission impacts and there are other values (ranging from 25- 100) people have used, we chose this because it is a conservative value and common way of interpreting methane emissions in the context of carbon dioxide emissions. Given the disparity in the mass of carbon in CO2 and CH4 emissions, simply adding them together can be misleading and often just reflects the pattern in CO2. We have added citations of other papers that similarly present CH4 and CO2 budget comparisons using CO2-equivalents.

RE-REVIEW COMMENT:

My point was not to question the numerical value of the GWP factor. Rather, the point was to highlight the fact that, no matter how widely it is used (I have done it myself…), the CO2-eq concept is poorly applicable to natural ecosystems as their continuous GHG exchanges do not induce radiative forcing or a warming/cooling effect. However, this is how the CO2-eq flux is commonly interpreted, especially when calculated with fluxes of opposite direction. Moreover, I do not understand how "the disparity in the mass of carbon in CO2 and CH4 emissions …" explains the use of CO2-eq. Misleading in what sense? If we are addressing the carbon cycle, then it is most logical to calculate the sum as CO2-C + CH4-C, but I would not convert CO2-eq to C and call the result carbon budget.

We respectfully disagree on the first point. These natural systems do induce a net warming effect, as human-caused climate change thaws permafrost, changes hydrology, and leads to other indirect effects resulting in the Arctic becoming a net source of carbon instead of the historical sink it has been. Discussing the sum of CO2 and CH4 as a mass of C is misleading because the mass of C from CH4 emissions is often orders of magnitude less than that of CO2 emissions, and thus when examining trends CH4-C + CO2-C ~ CO2-C. We have updated the main text to specify "CO2-eq carbon budgets" in place of "carbon budgets" whenever we are referring to joint CO2 and CH4 scaled budget results.

Example see Lines 420: We can combine the posterior distributions of scaled carbon from all three footprint model results to calculate a single CO2-eq carbon budget estimate that accounts for across-model uncertainties.

(4)

REVIEW COMMENT - Table 1: It would be useful to show similar percentages for the average footprint-weighted land cover proportions during the study period, estimated with different footprint models.

AUTHORS' RESPONSE - The full distributions for the footprint-weighted landcover proportions for each of the three footprint models are in Figure 2 and in the SI. We chose to display the tower area proportions instead of footprint-weighted averages since none of the distributions are Gaussian and displaying the average is not a useful metric or central-tendency in this case. We believe that including the actual distributions where the range and variance are clear is a better way to communicate this.

RE-REVIEW COMMENT:

Consider a flux (or any surface parameter) q_i that is constant over a certain period of time; i denotes the land cover class. It is easy to show that the mean flux observed by EC during this period can be expressed as sum_i(<f_i> q_i), where f_i is the 30-min footprint-weighted areal proportion of the land cover class i and <.> denotes temporal averaging. Thus, <f_i> indicates the relative contribution of each land cover class to the mean GHG flux or balance during the averaging period. I find this a useful metric that does not depend on the distribution of the 30-min f_i data.

That is an interesting comment and we will consider it in future work.

(5)

REVIEW COMMENT - line 127: What criterion was used for nonstationarity?

AUTHORS' RESPONSE - We used the Foken et al. 2004 method, as is cited in line 126 (implemented in EddyPro). This method assigns integers to indicate fluxes that pass or fail these QA/QC checks on the high frequency data. This method is a widely used in established EC networks (e.g. Ameriflux and Fluxnet).

RE-REVIEW COMMENT:

Yes, but the question was about the value adopted for the relative non-stationarity parameter. If you wish to refer to the QA/QC flagging system, then please specify in the manuscript how this was applied.

We have clarified that we removed bad data flags (value =2) from the overall flagging described in the updated citation.

Lines 125-126: We removed fluxes with nonstationarity (flags = 2 in the overall flag system) (Mauder and Foken 2015).

(6)

REVIEW COMMENT - line 395: Why would increasing the number of model parameters increase uncertainty? I would have assumed that this makes the model more flexible and thus decreases uncertainty.

AUTHORS' RESPONSE - Increasing uncertainty with increasing complexity is a common occurrence across modeling. See Puy et al. 2022 (https://www.science.org/doi/10.1126/sciadv.abn9450) for a nice summary.

RE-REVIEW COMMENT:

The Puy et al. (2022) paper deals with ”process-based mathematical models that do not (or cannot) rely on a training and/or validation dataset”. You do not use a process-based model and do have a training/validation dataset. While increasing uncertainty with increasing complexity may be common, it is not universal, hence my original comment.

In our study, because we are carrying through all of the uncertainty associated with every parameter into the regional carbon budgets, when we estimate twice as many parameters in more complex models we have more sources contributing their uncertainties. This section is explicitly about uncertainty in scaled carbon budgets, a derived quantity, not uncertainty around predicted values in the validation dataset, which is indeed often lower in the complex model versions. We have added text to clarify.

Lines 399-400: For most months the complex map solutions were slightly more uncertain, a consequence of estimating almost twice as many parameters and carrying through all of their uncertainties.