Author response to interactive comment CC1 submitted by Cole Brachman on Jul 20, 2021

In the document below, the reviewer comments have been copied from the original review and are shown in black font, while the author comments have been added in blue.

The manuscript aims to determine the role of grazing in carbon cycling through CO2 and CH4 gaseous fluxes in wet tundra habitat by the means of the large-scale herbivore reintroduction experiment of Pleistocene Park. The authors measured ecosystem respiration (Reco), Net Ecosystem Exchange (NEE) and CH4 using chamber methods and a flow through gas analyzer over seventeen days in five different plots distributed over two sites, one for the grazed (GR) condition within Pleistocene Park and one for the ungrazed (UGR) condition located nearby to the park. Gross Primary Productivity (GPP) was also calculated from Reco and NEE. The fluxes were interpolated based on the chamber measurements, air and soil temperatures, and soil moisture conditions over the measurement period. There were differences in the fluxes between the site conditions, which were primarily attributed to grazing having a drying effect on the GR sites. These initial findings, if further verified with additional measurements as outlined below, could result in some important implications for the role of grazers on the tundra landscape. Overall, this paper hints at some very interesting connections between carbon cycling, environmental conditions, and grazers but require some additional measurements to support the bold claims as they are currently in the manuscript.

Major comments:

The data are not enough to support the claims being made in the manuscript. The limited number of independent measurements and an unequal sampling design undermine the conclusions reached about the relationships. 17 days of measurements give an accurate estimate of the fluxes over that period, but do not necessarily represent the whole growing season. It is mentioned in the paper that these should be treated as a snapshot in time (especially for the GR plots), however, I do not believe the main takeaway points as they are written are properly taking that caveat into account which can result in some miscommunication of the strength of the findings. Additionally, only having two plots in the UGR condition, and only measuring those plots four times (4 days compared to 9 days for the three GR plots) makes accurate comparisons between the treatment types difficult for the full measurement period.

We are aware that, based on the limited available database, particularly quantitative results are associated with considerable uncertainties, but we are confident that this fact is well reflected in the discussion of the material. To further emphasize the limited database and temporal coverage, the title will be modified in the revised manuscript version, now reading "Grazing enhances carbon cycling, but reduces methane emission during peak growing season in the Siberian Pleistocene Park tundra site". Also, a new statement will be added to the end of the abstract (see comments to reviewer 1).

At the same time, we are certain that our results capture the dominating qualitative shift in ecosystem characteristics and carbon cycle dynamics that follow a decade-long, intensive grazing disturbance in these very sensitive Arctic wetlands. Even though our carbon flux estimates cannot be proven to be representative for larger areas outside of the flux footprint, we believe that our study provides valuable and novel insights into the impact of such management practices, and their application as a potential tool to protect Arctic permafrost from degradation under climate change.

Many studies covering novel, uncharted scientific territory in regard to method and/or location may be associated with a larger uncertainty compared to repeating established methods at previously studied locations. While we do not intend to discount the scientific contribution and merit of the latter, it may be rather incremental. From all possible forms of scientific inquiry, our abductive method is more speculative, but we strove to provide and include all information at our disposal in support of our results and claims. We will further strengthen this aspect by adding a more detailed discussion of the shortcomings, as also documented in our responses to the reviewer comments.

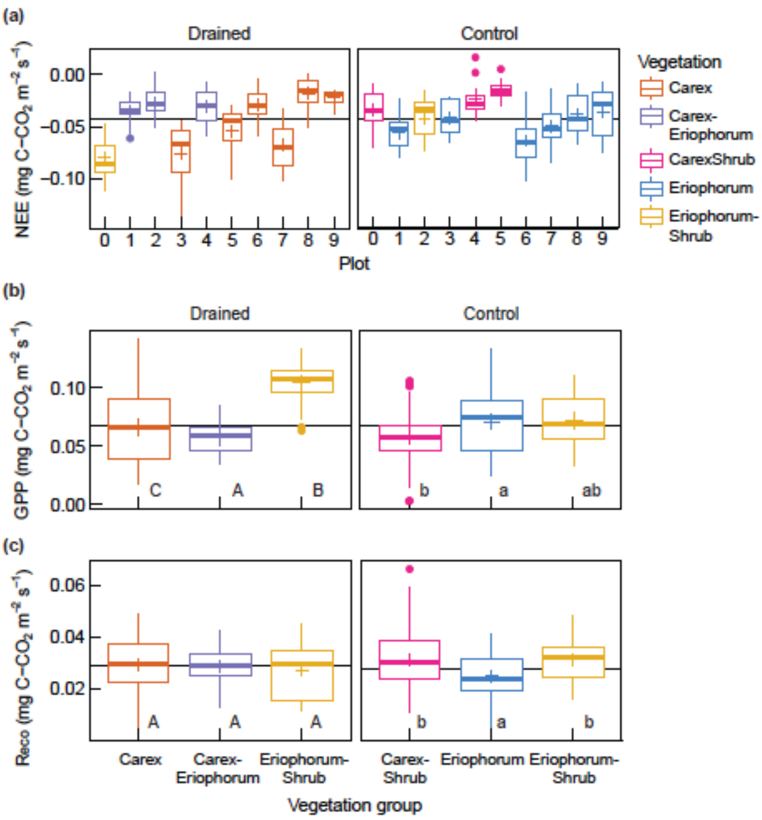
The two selected UGR plots had large differences in their GPP and NEE measurements and may not be a good representation of these sites. Selecting additional plots from the 10 previously established UGR plots for measurements would help to more accurately determine average flux values. The individual UGR plots are also showing very similar fluxes as the GR plots, but not consistently (see table 3). For instance, UGR 1 have similar GPP and CH4 as the GR plots, while UGR 2 seems to bring down the average GPP in the UGR plots. In addition, the UGR plots were not measured on the same days. This clearly demonstrates how the low replications undermine their conclusions.

We agree with the reviewer that a database comprising only two sampling sites cannot provide a statistically sound representation of an observation site featuring fine-scale variability. However, it was clear from the onset of the experiment that a data coverage of just 2.5 weeks could not provide a comprehensive assessment of grazing impacts. This was never our intention, as mentioned already above, and this is also clearly stated in the manuscript text.

The rationale behind our site selection was already discussed and explained at length in our responses to the comments of Reviewer 1, these statements are therefore repeated here:

The decision to work with only two reference sites (UGR) plots at the Ambolikha site was based on practical considerations. In principle, we could have used up to 10 sampling locations which had been established in earlier experiments. However, plots were spaced 25m apart, meaning that the observation system has to be moved between sites when switching locations, as opposed to the GR sites, which were co-located within a narrow radius. Spending time for moving the system implies less time for actual measurements, which is why we wanted to reduce it. Therefore, the choice was made to only sample two sites.

We realized that the description in the submitted version of the manuscript chosen to justify the UGR site selection was somewhat misleading. While fluxes at the two selected were actually indeed close to the mean fluxes across the transect, our choice was rather motivated by the ecosystem structure. While we cannot give more precise information on the GR sites before grazing started, we know from personal communication that the managed area used to be a waterlogged tussock tundra. Out of the 10 plots that were available at the UGR site, six are dominated by cotton grasses (Eriophorum), with few or no tussocks present (see Figure 8 from Kwon et al., 2016, copied below). Two more sites (IDs 4 and 5) were placed on a small ridge, and were therefore significantly drier, and dominated by shrubs. We therefore selected the only two locations, IDs 0 and 2 in the control section, featuring the desired vegetation structure for investigating the effects of grazing. Studies with a different scope may have enabled a random site selection to improve estimates of uncertainty due to site-specific bias.



The authors actually cannot follow the rationale that fluxes across chambers need to be measured on the same measurement days in order to be comparable. This is clearly not practical when sampling sites are located far apart, and even impossible for experiments that include large numbers of sampling spots that are regularly revisited. In our setup in the Chersky region, taking instrumentation from the grazed to the ungrazed study site, or vice versa, would have taken about two hours – precious time that we preferred to rather invest into actual measurements. As long as there are no systematic and fundamental differences in weather conditions between measurement days, we concluded it is fully sufficient to aim at capturing fluxes across a wide range of environmental conditions at each site in order to allow fitting response functions.

Site differences between the GR and UGR plots make it difficult to determine if the differences in fluxes are actually due to grazing effects and not moisture itself. Stronger evidence of the GR plots being water-logged throughout the growing season ~30 years previous, and that the drying of the site is due to grazing, is necessary to solidify the link between grazers and fluxes. Alternatively, flux measurements on wetter areas in Pleistocene park, and dryer areas in the UGR site may help disentangle the effect of moisture from the effects of grazing.

We would have liked to include some data-based evidence in the manuscript that demonstrates that both sites had similar pre-treatment characteristics, and only started to diverge with increasing grazing pressure at GR over the past decades. However, direct measurements of ecosystem characteristics within Pleistocene Park from the 1990s or before are not available, including soil moisture assessments that could help to compare soil hydrology over the past decade in connection with the grazing management.

Also remote sensing products such as e.g. LandSat time series, which are available in several scenes per year since 2000 for both study areas, turned out to be ill-suited for this particular purpose. As we obviously lack in-situ observations from the pre-treatment stage, we resorted to discussing this aspect thoroughly in the manuscript while mentioning that potential differences in pre-treatment conditions may add a systematic bias to the differences in carbon fluxes obtained from our chamber measurements. As already mentioned in our response to Reviewer 1, in the revised version of the manuscript we will further tone down some statements in the abstract, and add an additional statement to the end of the abstract:

"Our results indicate that grazing of large herbivores may promote topsoil warming and drying, this way effectively accelerating CO_2 turnover while decreasing methane emissions. Lacking quantitative information on the pre-treatment status of the grazed ecosystem, however, these findings need to be considered as qualitative trends, while absolute differences between treatments are subject to elevated uncertainty. Moveover, our experiment did not include autumn and winter fluxes, and thus no inferences can be made for the annual NEE and CH_4 budgets at tundra ecosystems."

Minor comments:

L 21: "Based on expert assessment", please delete.

The quote 'expert assessment' was actually taken over from the Schuur et al. (2015) reference quoted in this sentence. However, we agree that this statement may be misleading, and therefore re-formulated to "Based on several independent approaches, it is estimated that 130 to 160 Gt C could be released by 2100 .."

L 53: The drawbacks of measuring fluxes only in the growing season were mentioned, however, this study also only measured fluxes during a subset of the growing season. Consider leaving this to the discussion section as the reader expects some mention of a whole-year upscaling when it is mentioned early on in the introduction.

This reference was also criticized by the other reviewers, and is addressed in more detail there. In short, we agree that it may be misleading to refer to year-round fluxes in the introduction when our study does not deal with them, and therefore removed this sentence.

In the introduction, there are multiple mentions of shrubs and the effect of shrubs on C dynamics (possibly due to a large amount of the reference studies coming from Scandinavia and focusing on reindeer browsing), but your sites are dominated by graminoids. I would suggest reframing the introduction to focus more on the effect of graminoids on C dynamics and their interaction with large herbivores. This is also not much elaborated in the discussion, and the introduction as it reads now give the wrong expectations on the manuscript.

We agree with the reviewer that shrubs are not a dominating factor for our experiment carried out in the Kolyma lowland region, though shrubs certainly are an important element for the vegetation composition within the floodplain. However, the term 'shrub' is mentioned exactly three times in the introduction: once in a general section on Arctic climate change that is not focusing on grazing, a second time when citing potential influences of herbivore grazing on tundra vegetation, and a third time when listing hypotheses postulated for the Pleistocene Park experiment. The most important of these statements, i.e. the second one, is directly followed by the sentence "Grazing has been shown to promote certain Carex species that produce a high belowground biomass, ..". We believe our use of the term 'shrub does not raise incorrect assumptions or expectations' in the reader and thus is not misleading.

In the discussion, Section 4.4 which focuses on "Grazing Impacts on Vegetation" actually strongly focuses on graminoid species, and their relationship to grazing. We therefore disagree also with the claim that graminoid interaction with herbivores is not much elaborated on. To cite some examples:

- almost all sedge-tussocks were in a state of decay, or had disappeared almost completely. In place of them or between their remnants, many single plant tillers (mainly Carex spec. and Calamagrostis langsdorfii) grew.
- the transformation from tussocks to grass mats by grazing, accompanied by a strong increase in belowground biomass, was already observed for montane biomes
- Some sedges found in Arctic environments, such as Carex aquatilis, were shown to benefit from muskox-grazing, since they feature strong root production and the ability to produce dense grass tillers, and therefore more easily recover from grazing
- Accelerated urea-nutrient uptake by living plants has been reported for upland tundra (Barthelemy et al., 2018), where graminoids were more efficient in using these resources compared to shrubs.

Suggest renaming the plots from grazed (GR) and ungrazed (UGR) to heavily grazed (HGR) and ambient grazed (AGR), respectively, unless there are no populations of grazing herbivores on the landscape at the ambient site (no information provided).

We already changed the site description accordingly, based on a comment by Reviewer 1: "Our undisturbed reference ecosystem is a tussock tundra site situated about 15km south of Chersky on the floodplain of the Kolyma River. Due to the very low natural abundance of grazing herbivores in the region, the influence of grazing disturbance on this dataset can be considered negligible.". We will therefore stick to the site descriptions GR and UGR.

L 200: Mann-Whitney U tests were brought up in the statistics section but I could not find the results or a figure on these tests. Since these measurements also are repeated measurements, you need to provide evidence that they are independent between days (your statistical unit) or perform statistical test considering the repeated measures.

We decided to adjust the statistical tests used for this purpose. In the revised manuscript version, instead of Mann-Whitney tests, a repeated measures ANOVA will be conducted prior to post-hoc pairwise t-tests and a correction of p-values with the Holm–Bonferroni method. In this way, we will provide evidence that the measurements are independent between days, and correct the results of the pairwise comparisons for multiplicity problems. This approach will be used for fluxes, soil temperatures and radiation measurements.

L 306-311: Coefficients of Variance (CV) were discussed to determine if the heterogeneity between plots were in an acceptable range. However, when compared to the paper cited as a reference for this metric (Davidson et al. 2002), the present study has half the number of total plots they are assessing over which could be a factor in the low values found. The Davidson et al. (2002) paper also suggests a formula for determining the number of measurements needed to ensure a decent variance around the mean, which could be a useful way to determine if the number of measurements taken are representative or if more measurements are needed. In addition, it is unclear what measurements the CV is calculated on. It should be the daily data, 4 measurements for UGR and 9 for GR.

We will add an additional sentence to this paragraph to highlight the fact that the database was not equally distributed between GR and UGR sites: "However, one has to keep in mind that the GR sites feature more plots than UGR, and also a higher number of observations, both of which may influence a comparison of derived CVs."

Equation 3, which corresponds to interpolating Reco from UGR plots according to section 4.2 (lines 320-322), includes the data from GR-3. The interpretation of data from the GR plots therefore differ from each other, and GR-3 is interpolated more accurately with the same formula as that for the UGR plots. This was mentioned on line 326 stating that the measurements are not representative across the GR plots, which poses problems for the final conclusions drawn regarding these plots.

Equation 3 was indeed used to interpolate Reco for both the 2 UGR sites and the GR-3 site. This is stated in lines 326f in the Discussion paper: "For that reason, at GR-3 also Tair was used to interpolate Reco, since …". Our interpretation of the fact that we find different response functions for Reco across the GR sites is that there is obviously some micro-scale variability within a seemingly homogeneous ecosystem, and that we were able to capture this variability through our three sampling sites. GR-3 appears to be slightly drier than the other 2 GR sites, which is also reflected in the CH_4 flux rates. However, we do not see how this poses a problem for the conclusions drawn in our study.

L 364-373: Is it possible to tie these vegetation changes into the differences in measured fluxes more directly? Maybe a reference on fluxes from tussocks vs. grass mats?

Unfortunately, we are not aware of reference studies that directly compare the flux rates between tussocks and grass mats under the same environmental conditions. Some of our own work on the Ambolikha site compared fluxes from tussock-dominated patches to those with dense cotton grass meadows, which is e.g. reflected in the figure from Kwon et al. (2016) copied above. Here, the cotton grass meadows (Eriophorum plots) featured higher GPP and lower Reco, compared to the Carex-dominated plots. However, this is not precisely a good reference for 'grass mats' that may develop under grazing pressure.

L 373-375: Were the addition of CO2 and CH4 from grazers themselves factored into any calculation of total fluxes from the sites?

No, direct emissions from herbivores were not considered in our estimates.

Clarification of the prevalence of these wet tussock tundra sites within and outside of Pleistocene Park would be a useful addition when visualizing how these results may affect the larger arctic region.

Current pan-Arctic vegetation maps are not yet detailed enough to differentiate wetland features such as e.g. wet tussock tundra. For example, in a recently published study by Olefeldt et al. (2021), Arctic wetlands were merely separated into 'permafrost wetlands' and 'permafrost bogs', and even this can be considered a big advance from aggregating all kinds of wetlands into a single vegetation class.

L 402: This sentence needs a reference at the end.

We will add the study by Göckede et al. (2017) as a reference here.

L 403: "only very inefficiently", consider revising.

The sentence we be re-written.

L 731 reference for Zimov et al. 2012, seems to have the incorrect initials for one author (F. S. Chapin).

This typo will be corrected in the revised manuscript.

References cited:

Göckede, M., Kittler, F., Kwon, M. J., Burjack, I., Heimann, M., Kolle, O., Zimov, N., and Zimov, S.: Shifted energy fluxes, increased Bowen ratios, and reduced thaw depths linked with drainage-induced changes in permafrost ecosystem structure, Cryosphere, 11, 2975-2996, 2017.

Kwon, M. J., Heimann, M., Kolle, O., Luus, K. A., Schuur, E. A. G., Zimov, N., Zimov, S. A., and Göckede, M.: 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, 4219-4235, 2016.

Olefeldt, D., Hovemyr, M., Kuhn, M. A., Bastviken, D., Bohn, T. J., Connolly, J., Crill, P., Euskirchen, E. S., Finkelstein, S. A., Genet, H., Grosse, G., Harris, L. I., Heffernan, L., Helbig, M., Hugelius, G., Hutchins, R., Juutinen, S., Lara, M. J., Malhotra, A., Manies, K., McGuire, A. D., Natali, S. M., O'Donnell, J. A., Parmentier, F. J. W., Räsänen, A., Schädel, C., Sonnentag, O., Strack, M., Tank, S. E., Treat, C., Varner, R. K., Virtanen, T., Warren, R. K., and Watts, J. D.: The Boreal–Arctic Wetland and Lake Dataset (BAWLD), Earth Syst. Sci. Data, 13, 5127-5149, 2021.