Comment on bg-2021-110
Cole Brachmann

The manuscript aims to determine the role of grazing in carbon cycling through CO$_2$ and CH$_4$ gaseous fluxes in wet tundra habitat by the means of the large-scale herbivore reintroduction experiment of Pleistocene Park. The authors measured ecosystem respiration ($R_{\text{eco}}$), Net Ecosystem Exchange (NEE) and CH$_4$ 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 $R_{\text{eco}}$ 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.
- 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.

- 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.

Minor comments:

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

- L 402: This sentence needs a reference at the end.
- L 403: “only very inefficiently”, consider revising.
- L 731 reference for Zimov et al. 2012, seems to have the incorrect initials for one author (F. S. Chapin).