Biogeosciences Discussions

Biogeosciences Discuss., referee comment RC1
https://doi.org/10.5194/bg-2021-110-RC1, 2021

Comment on bg-2021-110

Anonymous Referee #1

Referee comment on "Grazing enhances carbon cycling but reduces methane emission during peak growing season in the Siberian Pleistocene Park tundra site" by Wolfgang Fischer et al., Biogeosciences Discuss., https://doi.org/10.5194/bg-2021-110-RC1, 2021

General comments

The authors describe the effects of a manipulation of grazing herbivores density in a continuous permafrost tundra ecosystem on carbon fluxes during the growing season and some of its predictors. The core data is an extensive set of NEE and R_{eco} measurements obtained over two weeks at peak growing season over three replicates of the high grazing density system and two replicates of the low grazing density system, and is accompanied by meteorological variables. Overall, gross fluxes GPP and R_{eco} were increased in the high grazing density plots, concomitant with an increase air and soil temperature and a decreased soil moisture content, while NEE was largely unaffected. CH_4 fluxes were lower in the high grazing density plots but with high variability between plots. The flux measurement dataset is valuable, and my main concerns lie in the choice of an unbalanced design which hampers statistical evaluation of the results, and that too few details are provided to justify the fact that initial conditions were comparable and throughout the Methods section. I would therefore recommend that these aspects be thoroughly improved before publication and try to provide suggestions for such improvements.

I appreciate that, since initial submission, further information has been added regarding the choice of the two replicates in the UGR site, but still think that beyond the choice of these particular plots, the decision of having only two control plots should at least be better motivated and the limitations it implies better discussed. Overall, that does not entirely appease my concerns regarding the unbalanced study design. On the contrary, by choosing two sites that are close to the average within that transect, rather than e.g. at random, the mean value is preserved but the variance is artificially deflated, possibly biasing comparisons. While this may or may not affect R_{eco} dynamics so much, as the differences are marked and variability seems limited, the differences observed in other variables and in particular CH_4 fluxes could be artefactual for this reason. In addition, UGR-2 is described as standing out on several aspects, from the lower soil temperature to the higher time lag of soil vs air temperature, to having only half as many flux measurements as the other plots (so that there is, in total, twice as much flux data for GR than UGR plots), which explains in part the poor measured vs modelled fit of GPP for

UGR-2.
I do not see a simple way to solve this issue, but perhaps simulating data based on flux measurements from Kwon et al 2016, using the relationship between the fluxes in the current study and those in Kwon et al 2016, could allow to carry out a sensitivity analysis to determine whether the choice of these 2 replicates affected the findings.

Comparing the "grazed" and "ungrazed" treatments is central to this manuscript, but their identity prior to the experiment is unclear from the information currently provided. For instance, as it is now GR sites are described as a wet lowland tundra that gets flooded every year, while UGR sites are a wet tussock tundra floodplain, it is unclear whether the distinction between the two is intended to avoid repetition or to convey a more fundamental difference between the sites. Beyond the more detailed comments below requesting that more detailed data on initial conditions should be presented if available, I would suggest reorganizing the part of the Methods section where the sites are described. In addition to the general geographic and climatic information about the area, it could be easier to follow if the authors would first describe the similarities between the two sites prior to manipulation (e.g. flooding, vegetation, etc.) before delving into what makes them distinct.

Specific comments

L30: Might be worth citing the recent review by Mekkonen et al 2021 found here: https://doi.org/10.1088/1748-9326/abf28b
L53-56: It's not clear how Pleistocene Park and the measurements presented here address non-growing season carbon cycling, in fact the abstract (L14-15) explicitly states that those are not addressed. I suggest reformulating or removing the reference to non-growing season fluxes from this section altogether.
L58-63: It's perhaps more a question of personal taste, but I find it a bit confusing to present hypotheses that are not tested in the current study (e.g. above-belowground partitioning of GPP, decrease in respiration from colder permafrost in the winter). I would think these wider considerations about the aims of Pleistocene Park itself would fit better in the discussion, or presented differently than a list of hypotheses near the end of the introduction, as that might give the reader the wrong impression that those are the hypotheses addressed in this study.
L83-86: Can you provide more precise information on that site, such as for how long the density of grazing herbivores has been increased, and by how much compared to the "ungrazed" site which arguably hosts large grazing herbivores in lower densities such as the rest of the Arctic, unless the site is too wet and not visited by grazers?
L91-93 and 101-103: Can you provide more precise information on e.g. vegetation composition and soil conditions prior to the onset of manipulating density of grazing herbivores? Something similar to Table 2 in Kwon et al 2016 would be a good start to support the central assumption that the sites were initially similar.
L103-104: For future reference, it would be good to mention to which plot numbers in Kwon et al 2016 the UGR-1 and UGR-2 sites in this study refer to. 2-0 and 2-2 are mentioned further in the discussion (L300) and should be mentioned here instead, but do not appear directly relatable to the denomination in Kwon et al 2016: does that mean Control-0 and Control-2?
L108-112: Please clarify how many probes were used and where. From the phrasing I would have expected one probe per plot for temperature, and three probes per site for moisture, but from Fig.1 and the corresponding Results section it seems like there was

only one probe in one of the GR plots and two in UGR. Please also clarify what was considered 0cm depth in the water-logged tussock tundra (e.g. water-table, top of the tussocks, between tussocks), I assume it is the soil surface between tussocks but I shouldn't have to make assumptions in the Methods section. Please also mention the logging interval and the procedure used for producing the interpolated data presented in Fig.1.
L113-115: Please clarify what is meant by plausibility limits, which offsets were corrected and how. Please also provide further detail on the interpolation of soil temperature data: which variables other than air temperature and incoming radiation were used, and how. In addition, indicating the date and value of individual measurements in Fig. 1 would be helpful.
L123-124: Please clarify whether a single relationship was used (if so, which one) or whether incoming radiation-PAR relationships were adjusted by incoming radiation classes.
L139-140: I assume that not all measurements were 2 minutes long since that time is mentioned as a maximum. I would expect an arbitrary threshold in change in CO_2 concentration was used to limit non-linearities from excessive CO_2 buildup or uptake within the chamber, but could not find that value here or in Kwon et al 2016. Could you please clarify that part?
L149-152: Is it correct to assume that these slopes were linear fits? If so, perhaps mention it at L150 instead of "steady", if not please clarify. In addition, a duration criteria must have been used for selecting the periods with a steady increase, please mention how long these periods had to be in order to be considered.
L152-155: I assume this is the reason for the uneven number of "utilizable" measurements in the different plots, if so please move the reference to Table 1 after this section instead of at L147. Please clarify in Table 1 what "utilizable" refers to. Please also mention somewhere why so few measurements were utilizable in UGR-2.
Appendix A: In line with my earlier comment about the number of probes used and their location, please clarify where the moisture / temperature data used for modelling comes from. Were GR-1, GR-2 and GR-3 models based on the same temperature data from the probe in GR-1? If so, was the data averaged between UGR-1 and UGR-2 or did these plots benefit from a distinct processing where they each had their own supporting data for the modelling?
L200: One of the assumptions of the Mann-Whitney test is that the observations are independent, but for both T_s and fluxes the data are time-series and are therefore not independent.
L199-201 and throughout the text: The statistical tests results mostly present *P* values. I could not find precise guidelines of in-house rules regarding the presentation of statistical results in Biogeosciences, but I would suggest the test statistic to be presented as well, and the degrees of freedom. If this burdens the text too much, please consider a supplementary table.
L201: default t-test may not be appropriate with unequal variances, particularly so with unequal sample sizes as well. Please mention whether the assumptions for using a t-test were checked and met.
L207 and throughout the text: When presenting mean values, please provide associated uncertainties in the form of SD, SEM or CI.
L223-224: Is there any data on this difference in air temperatures before the onset of the experiment? I.e. does this reflect initial differences or an effect of altered vegetation and soil conditions?
L229-234: There is no mention in the Methods section of how these values were computed and it is unclear what the *P* values presented refer to, please clarify.
L236: "clearly" is strong phrasing considering the absence of replication. Given the 100h lag at 15cm depth, effects of the change in weather pattern should only be visible in the last couple of days, at best, but one would expect deeper layers to have higher thermal inertia and therefore not to see an effect of the change in overall weather pattern by the end of the study period. In that respect, I do not think it is justified to carry out separate

tests for the two weeks at this depth.
L249-248: There is no mention in the Methods section of how or when thaw depths were measured or inferred from temperature data, please correct this.
L252-277 and Figure 2: Judging by L199-200, the pairwise comparisons presented as letters in Figure 2 were computed by running 10 different Mann-Whitney tests per variable, plus one for the averaged values. If that is not the case, please describe this in the Statistics section, if that is the case, please clarify it in the Statistics section as well. In both cases, please indicate (how) were the *P* values adjusted for multiple comparisons. Beyond concerns about the Mann-Whitney test assumption of independence of observations expressed above, I would advise running an omnibus test prior to post-hoc pairwise comparisons. With a balanced design, a repeated-measures ANOVA could be a correct way to account for dependent observations within a plot. Considering the central role of flux data in this manuscript, their statistical treatment should be improved.
L256-258: It is unclear what "flux rates" refers to in the first part of this sentence: NEE, R_{eco}?
L286-287: When were the collars installed at the GR site? Presumably after setting up the wooden fences to prevent trampling by the herbivores, but please mention this here or in the Methods section.
L293-295: Why not mention tussock-forming plants here? As far as small-scale heterogeneity is concerned it seems odd not to mention one of the main ecosystem engineers of these systems.
L308-309: It might be good to remind here that when comparing CVs of GR and UGR one should keep in mind GR having 50% more plots and ~100% more measurement points.
L324: See above at L108-112, this is not a reminder and this information should be stated more explicitly in the Methods section.
L329: I assume "the actual measured values" refer to R_{eco}, but please clarify.
L351-355: It would be good to mention examples of which such operations might be confounded with the effects attributed to the increased grazing herbivore density.
L357-359: See my earlier comment about L91-93 101-103, in absence of more detailed data such photographs may be an interesting supplementary display item.
L394: "mostly likely" should be "most likely", but the phrasing is a bit strong for an hypothetical future development, which to date is in contradiction with the observations as shown in Fig 1. While I understand the hypothesis of a cooling of the soil and grazing-induced protection of permafrost in Pleistocene Park, it is hard to ignore that Fig 1 shows an almost twice as deep active layer thickness in the grazed site. Either the hypothesis is correct but the sites differed drastically in active layer thickness prior to the experiment, or the effects observed after 22 years of manipulation contradict the expected consequences of the hypothesis. A transient regime is possible but less parsimonious, and "most probably" or "most likely" are too strong for that to my taste.
L404-406: Liquid water has a fairly high thermal conductivity, a comparison between values for a compacted soil and a water-logged soil could be useful information here.
L408-410: It is hard to say for 25cm depth since the data is not shown, but for 35cm it would be good to remind that the observed difference in soil temperature is lower or similar to the observed difference in air temperature.
L414-415: In line with the previous comment, this sentence could be complemented by starting it with "Barring differences prior to the onset of the experiment".
L416-419: Considering that no difference was observed in growing season NEE and that in presence of grazers, a larger fraction of NPP is removed by herbivory, this argument should be substantiated with above- and below-ground plant biomass measurements or a complete C budget. In their absence, it is speculative and because this is not central to the reasoning, I would suggest removing it.
L419-420: See above at L394, this is speculative and in direct contradiction with data presented in Fig. 1.
L453 and 456-457: This is speculative, please use less strong phrasing.

Technical corrections

L27: change sentence order
L57: facilitates -> allows
L116-119: I would suggest using GR and UGR rather than Pleistocene Park and Ambolikha for consistency.
L165-166 and L200: Rstudio is only a GUI software to R and does not do calculations. Please move the mention to the software used to the end of the statistics section, and provide adequate reference including R, the version number and the appropriate citation (e.g. R Core Team. R: A language and Environment for Statistical Computing. (2021)
L200: "(?)"?
L226 – Figure 1: I would recommend making two separate panels out of panel (a). I do not think the y-axis break simplifies the figure, and the factor 5 change in axis scale would be more obvious that way.
L437-439: Please consider rephrasing, the current syntax poses "increases in primary productivity" as an explanation for increased GPP.
L480: "differences in NEE were not pronounced" -> "no differences in NEE were found"

Appendix A, L494-496: This sentence would be easier to understand if the information was split across several sentences, please rephrase it.

Review criteria:

Does the paper address relevant scientific questions within the scope of BG?
Yes

Does the paper present novel concepts, ideas, tools, or data?
Yes, effects of a long-term herbivory experiment on C-cycling in a continuous permafrost system

Are substantial conclusions reached?
Somewhat

Are the scientific methods and assumptions valid and clearly outlined?
Methods and assumptions need to be substantiated

Are the results sufficient to support the interpretations and conclusions?
There is some amount of speculation not supported by the results, but not overly much

Is the description of experiments and calculations sufficiently complete and precise to allow their reproduction by fellow scientists (traceability of results)?
No, further work on the Methods is necessary

Do the authors give proper credit to related work and clearly indicate their own new/original contribution?
Yes

Does the title clearly reflect the contents of the paper?
Yes

Does the abstract provide a concise and complete summary?
Yes

Is the overall presentation well structured and clear?
Mostly, I suggested some changes

Is the language fluent and precise?
Some improvements can be made to the language, I suggested some but not all

Are mathematical formulae, symbols, abbreviations, and units correctly defined and used?
Yes

Should any parts of the paper (text, formulae, figures, tables) be clarified, reduced, combined, or eliminated?
Some sentences should be deleted, perhaps a paragraph in the discussion. Further details should be provided in the Methods, and the statistics should be better described both in the Methods section and in their presentation throughout the Results section. Fig.1 could be improved, legend of Fig.2 should clarify the source of post-hoc / significance letters.

Are the number and quality of references appropriate?
Yes

Is the amount and quality of supplementary material appropriate?
Yes