Dear Editor, dear Referees,

Here, we adapted the individual replies given earlier to fit the ‘combined opinions’ of referees and editor. We found that the comments helped to mend some flaws and recognize sections that were not sufficiently clear. As a result, we present a substantially improved manuscript. Where we did not agree with the reviewers’ points raised, we gave detailed explanations and arguments on our views. Please find our point by point replies below.

Please note that line numbers in the reviewers’ comments refer to the original manuscript; but the line numbers in our responses refer to the revised manuscript with track changes displayed.
For the ease of reading we quoted the comments by the editor and two anonymous referees with an indent first, followed by our responses with ‘**A >**’

> I would like to add that both reviewers expressed some concern about the frameing of the experiment as transplantation experiment, and I agree that the nature of the experiment should be clear already from the title. Please consider adding "... in an transplantation experiment along an altitudinal gradient" so something similar but more readible to the title.

A > Agreed. New, comprehensive title is:
‘Sub-alpine grassland productivity increased with warmer and drier conditions, but not with higher N-deposition, in an altitudinal transplantation experiment

> I would like to point out the data policy of Biogeosciences (https://www.biogeosciences.net/policies/data_policy.html). While it is of course not the intention to make the raw data of the experiment accessible, please consider seriously the request of reviewer #1 to make the statistical analysis as reproducible as possible

A > We will deposit the data necessary to reproduce the statistical analysis on DRYAD (https://datadryad.org/stash). The relevant pieces of R code will be part of the revised Appendix.

Anonymous Referee #1

Received and published: 13 October 2020

1 General comments

First, I would like to see more information about the plant species composition of the experimental monoliths. Qualitative results can be informative too. This could be a few sentences in the methods. Photographs might also be helpful.

A > We added more information on species directly in the new M&M l. 166-170. But to keep the MS as lean as possible we refer to the Wüst-Galley et al. (2020), which reported in detail on functional group responses in the same experiment. Please find a comment on photographs below.

Generally I would be slightly concerned about the inference obtainable from the elevational gradient. If many factors change in a correlated fashion along the gradient, such as temperature, moisture, and historical human/grazing pressure, it is hard to tell which factor is the driver.

A > Indeed, in a complex system with a high number of interacting environmental factors, there is an equally high number of drivers. For this very reason, we have termed the sites at different altitudes 'Climate Scenario', which includes all what the reviewer claims. To uncouple altitudinal effects from soil moisture (and nitrogen) effects, we have set up an irrigation and N deposition treatment in a factorial design. To this respect we found that soil moisture and temperature, resulting from the Climate Scenario (CS) site at a specific altitude, both drive the plant productivity response (cf. Fig. 2 a,b), but the N deposition does not in a significant way (cf. *P* values in Tab. 3).

The management history of the sites of origin is very similar, but in concert with the strong edaphic factors, an effect on the present plant communities cannot be excluded. We regard this element of heterogeneity as an advantage, as it is a factor that supports the general applicability of our results.

I understand the limitations of the design and I don't think it's necessarily a flaw, but this is something that should be addressed more openly. This is also the case when discussing how soil moisture integrates information on both temperature and moisture; this could also be viewed as confounding the effects of temperature and moisture.

A > Agreed, the moisture of a Climate Scenario is not independent from the temperature. As stated above, we have termed the sites at different altitudes 'Climate Scenario' and analyzed the data accordingly. We assumed that the unavoidable temperature × moisture interaction closely resembles true climate change conditions, much better than an experimental manipulation of temperature or moisture alone would do.

Indeed, both moisture and temperature were related to productivity in a concerted way (cf. Fig. 2, Appendix Tables A4 and A5). From these analyses, it can be deduced that moisture was the stronger determinant of productivity, while extreme conditions of either predictor restricted plant growth.

One other point I would like to raise about the inference is that the warming treatment is confounded with site of origin. For example, the communities subjected to highest warming were those that were moved from the highest elevation. Therefore it is difficult to say whether the different levels of warming, or the composition of plant and soil communities from each of the sites of origin, led to the different productivity responses. This should be addressed as well.

A > There seem to be a misunderstanding of our design. We wish to clarify that, we did not create a warming treatment by transplanting from origins of different temperature to a common site of uniform temperature. Instead, all sites of origin have very similar temperatures and altitudes. In contrast, the experimental site, where the turf monoliths are transplanted, contained 6 climate scenario sites along a c. 700 m altitudinal gradient. Thus, six different climate treatments were established.
This misunderstanding has been clarified in an earlier round of comment/response.

For reproducibility, please make the code and data available in a repository so that readers can reproduce the results of the statistical analysis. This is especially important for the mixed model specification. Sometimes it is difficult for the reader to determine the exact model specification from the verbal description but it is easier if they can see the code.

A > We will deposit the data necessary to reproduce the statistical analysis on DRYAD (https://datadryad.org/stash). The relevant pieces of the R-code will be part of the new Appendix.

2 Line-by-line comments

Line 10: The abstract does a good job of stating the results of the study but it does not do a good job of stating the motivation, novelty, or broader significance of the study from the outset. Please revise accordingly.

A > Helpful point. We upgraded the Abstract such that aspects of motivation and broader significance are included.

Line 55: The claim that multifactorial experiments necessarily will improve predictions is debatable. Please expand on the reasoning behind this claim.

A > The paragraph following our claim (new l. 104 ff) is dealing with the interpretation of multifactorial (or multilevel) vs. unifactorial experiments. It contains seven references to support the argument.

Line 76: The hypotheses need to have a little more justification or explicit statement of the reasoning why the particular directions of the effects and interactions are expected. For example, are there other studies that show similar effects or are the expectations derived from first principles?

A > Our hypotheses are not derived from first basic principles, as these are hard to gain due to the complexity of interacting climate change factors. Rather, our hypotheses relate to specific aspects that can be expected given the cited literature in the Introduction.
For example, we hypothesize that
'1) The effect of warming on plant growth would be beneficial at moderate warming levels, but detrimental at high warming levels.' (new l. 140-141).
Related to this hypothesis, the Introduction section mentions how warming at high altitude vs. lowlands reduces temperature growth limitations, rather than causing heat stress. Studies on the warming effect in cold environments (in part with inconsistent results) are given in detail (new l. 91-99).

Line 85: It is interesting that southerly exposed slopes were chosen for the study. They tend to be drier and warmer than slopes with different aspect at the same elevation. I would expect the plants living in these microclimates to be especially responsive to the warming treatment. Is this something worth briefly mentioning?

A > A uniformly southern exposure (and identical altitude), as opposed to different exposures, was chosen to minimize differences in climate (temperature, moisture, radiation) between sites of origin. Also, in this region of the Alps the majority of southerly exposed slopes is used as summer pastures. In contrast, the more the slopes are pointing away from the sun, the more likely they are forested and not suitable for our grassland research. Surely, one may assume that an adaptation to heat and drought has occurred in these plant communities, resulting in an improved tolerance against extreme events. This may make them more likely to outcompete plant communities from moist and cool habitats in the future. We do not see though, why plants from warm and dry habitats should be more responsive to warming, judged on a 'productivity-increase per warming' basis.

Line 106: A picture says 1000 words. It would be great to have some photos of the environment at the study sites, either as a main-text figure or as a supplement.

A > That is a hint we were waiting for ☺. We are happy to show the AlpGrass site and the landscape it is set in, and have added four photos to the Appendix.

Line 116: Similar to above, its would be nice to have a picture of the experimental setup.

A > Cf. above

Line 150: Is there a justification for the threshold for growing degree days being set at 0C? The same goes for the 40% soil volumetric water content threshold.

A > Indeed, the many degree day baselines for crops or other individual species are usually higher. We chose a 'generic' 0°C baseline, because - in a mountain environment - already low amounts of thermal energy play an important role. Furthermore, lacking a single target species to focus on, but working on multi-species communities instead, none of these 'specific ' baselines appeared more appropriate.

The SWC < 40 %-threshold does not imply plant growth limitation. Instead, it is an empirically developed contrast for differences in the soil moisture status between the CSs and between years. More time below the threshold simply means a 'drier period' in relative terms. This is also described in new l. 260-263.

Line 196 (statistical analysis section): I'm not sure I understand the reasoning behind assigning CS as a fixed effect but site of origin as a random effect. From my reading of the methods those are the same thing. Can you please clarify this?

A > CS represents the climate scenario treatment along the altitudinal gradient at the AlpGrass experimental site. It is an inherent treatment factor, similar to the irrigation and the N-deposition treatments. Because we wish to make specific tests and statements of the effect of the CS (or even one specific CS), this treatment has to be specified as a fixed effect. By contrast, 'site of origin' represents the grassland sites where the monoliths were excavated (and later transplanted to the AlpGrass experimental site). We do not investigate the effects of the origins and do not make any tests or statements on origins, therefore this variable is fitted as a random factor.

Line 216: Please include some details on the GAM fitting procedure, such as functional form of splines, etc. Were the defaults from the mgcv package used? If important inference is drawn from the GAMs, it would be good to assess the sensitivity of the results to choices made in the GAM fitting process. As written, it is not reproducible.

A > Indeed, we used the defaults from the mgcv package, which one exception. The 'gamma' statement of the gam() function has been increased slightly to increase the degree

of smoothing (to result in a smoother fitted line). This, however, did not (or only marginally) influence the inference and conclusions drawn from the model, i.e. *P* values for smooth terms reported in the main text and Tables A4 and A5 were highly significant in either case. To improve clarity, we provide the GAM specification in the Appendix.

> Line 277: Because all columns of Table 4, besides the two leftmost, are in the same units (mean and SE of aboveground biomass yield), it might be better to convey the information in this table with a figure. Currently it is difficult to visually extract the most salient patterns from the table. If you do not want to use a figure maybe another possibility would be to use colors or cell fills to show where the highest values in each year were recorded.

A > We visualized the response at the different CS in several panels (Fig. 1, Fig 2.), but for clarity reasons we did not split the data by years. Generally, we wish to put few emphasis on the within year results, as minor changes in productivity from year to year must be expected and are usually hard to explain. However, we are thankful for the idea of shading the cells with highest values in Table 4, and did so accordingly.

> Line 281: I am confused why -7.7% is described as an increase, is it a negative or positive change?

A > When viewed on our computer monitors l. 281 always says +7.7 % (‘plus7.7 %’), not -7.7%, in all versions of the document. Thus ‘increase’.
Maybe there is a *.pdf-file to printer communication problem, in case you worked on a printed copy?

> Line 289: Refer to the statistical test result (I am assuming this is Table A2?) that supports the statement that there was no significant interaction between N treatment and CS or irrigation.

A > Thanks for drawing attention to this. We added the reference (new l. 458). It is actually Table 3, where the non-significant single factor N treatment is reported, as originally referred to in the previous line.

> Line 316: "climate scenario warming" is a confusing phrase. Do you mean warming consistent with some particular climate scenario?

A > The sentence has been clarified (new l. 506-507).

> Line 390: I found this paragraph to be a little confusing. Are you referring to results from the present study or previous studies in the literature? Also, because you mention

specific species responses to N addition from other species, it would be more interesting if you would draw a more direct connection with the present study. Were there any individual species that you can point to their responses?

A > We improved the wording of the two related paragraphs. Now it should be clear that these paragraphs discuss N-effects in previous studies regarding whole plant community responses versus single key species responses (new l. 588-621).

Line 425: I am not sure what the grounds are for stating that subalpine grassland productivity will increase with warming. Is it necessarily the case that climatic conditions will "move up" in elevation – maybe there will be novel and unpredictable combinations of temperature and moisture not tested here.

A > The future climate may actually show ‘novel and unpredictable combinations … ‘. However, we did not speculate about unpredictable combinations of the future climate. In our experiment, we combined three key factors related to plant growth and measured soil moisture and thermal energy, all of this reflecting many possible combinations of future climate change factors in our tested environment. Given these treatment combinations and related information, we found clear indication that yields were increasing with increasing climate scenario mean temperatures (warming). In that context, we indeed assumed that climatic conditions will ‘move up’ under global warming conditions. This, however, is a most conservative and reasonable choice and keeps the number of necessary assumptions as low as possible.

Anonymous Referee #2

Received and published: 20 October 2020

Biogeosciences bg-2020-322: The rising productivity of alpine grassland under warming, drought, and N-deposition treatments

General Comments

In their manuscript titled “The rising productivity of alpine grassland under warming, drought, and, N-deposition treatments”, the others describe a novel experiment in which monoliths of soil and turf were transplanted across an elevational gradient combine with fertilization and water addition treatments. After four years of growth in the transplanted location, the others describe how plant productivity in the monoliths responded to the interaction of different temperatures (comparing climate at the transplant location to the original site where the turfs were harvested from), fertilization, and increased moisture, as well as the interactive effects of these three treatments. The results of this study showed that intermediate levels of warming increased plant productivity, even in drier conditions. Increasing the precipitation received by some monoliths had only marginal effects on plant productivity, while fertilizing the plots with nitrogen solutions had no discernable effect on plant productivity.

While this experiment is truly novel in its use of monolith transplants to simulate climate change in conjunction with two additional global change treatments in order to understand how multiple facets of global change will impact productivity, I have several concerns regarding the framing of these treatments, the metrics used to communicate and aggregate results, and the overall clarity of the manuscript. In particular, while transplanting monoliths to new elevations does of course impact climate, and in some cases results in warming, characterizing this experiment as a “warming experiment” is disingenuous. I encourage the authors to refer to their experiment as is, a transplant experiment across an elevational gradient. Furthermore, it is also a misnomer to refer to the precipitation manipulation component of this experiment as a “drought treatment”, as water was added to some monoliths instead of removing precipitation, as when using rain-out shelters etc., to simulate drought.

A > Indeed, in the headline we implied that we have a warming treatment, even though what we applied is an altitudinal transplantation treatment. This has been a commonly raised issue and we have altered the title accordingly.

We do not refer to the irrigation treatment as a ‘drought treatment’, but as ‘irrigation treatment’ (eg. section 2.3 in Material and Methods, new l. 265 ff). Drought, as a productivity limiting factor, is not a treatment per se in the experiment, but a consequence of the downward transplantation. The supplementary precipitation (not mentioned in the headline)

is a treatment to mitigate drought conditions. We feel that all of these aspects are appropriately addressed in the revised text.

My detailed line comments below elaborate on these concerns as well as my suggestions and critique of the metrics that the authors chose to describe climate in this study.

Line Comments

34–"... to have beneficial effects": Beneficial effects on what?

A > This has been improved (new l. 78-79).

35-36: Clarify what you mean by "initial water supply"... Water resources at the beginning of the growing season are generally plentiful? But this would be the case only for plants that emerge early in the growing season, i.e. depends on phenology of plant species.

A > Yes, water resources at the beginning of the season are generally plentiful. As we state in the following clause: 'because even a small winter snowpack supplies a large soil moisture resource in spring' (new l. 80).

Plants in subalpine grasslands are all perennial and usually start greening even before the snow-cover has completely disappeared. Within a given community, they reveal little phenological differences because the growing season is short; thus, growth and flowering peaks for the great majority of plants at the same time.

38–"kg N ha-1 a-1": These units are unconventional, instead of a-1 (per annum?) I typically see yr-1 when describing nitrogen deposition rates.

A > Yes, 'per annum'. Not unconventional. The SI convention for English year is 'a'.

45–"...showed a twofold productivity increase": In response to what treatment?

A > The sentence has been clarified. For the revised MS we will complement ' … *up to* twofold …' to better reflect the quoted author's statement (new l. 91-93).

47–"...grasses were favored over forbs and sedges by drought and warmth": This seems unclear, what do you mean by "favored by drought and warmth"? Productivity of forbs and sedges increases with warming and drought?

A > The sentence has been clarified. New sentence is more specific (new l. 93-96).

61–“...if only a short or linear segment out of a larger range of biologically possible responses is represented in the data.”: There is some indication that productivity relationships revealed in manipulative experiments actually encompasses even more variation than occurs naturally (see Jochum et al. 2020. Nature Ecology and Evolution).

A > Here we are making a point to encourage inclusion of many factors and factor levels in climate change experiments to avoid wrong interpretations by missing treatment combinations or by interpolating between data points.

With respect to biodiversity experiments, Jochum et al. (2020) found that biodiversity experiments ‘have greater variance in their compositional features than their real-world counterparts’. The authors later conclude that this does not impair the applicability of the results of biodiversity experiments: ‘ … our results demonstrate that the results of biodiversity experiments are largely insensitive to the exclusion of unrealistic communities and that the conclusions drawn from biodiversity experiments are generally robust’ (Jochum et al. 2020). We want to emphasize that our experiment does not study effects of or on biodiversity.

67–I think that I am still confused by what you mean by "factor levels"... Does this refer to consideration of multiple global change factors, or does it refer to the magnitude of the global change treatment imposed by the experiments?

A > It is the latter (“the magnitude of the global change treatment imposed by the experiments”). We have clarified that sentence.

68– “Here, we present four-years of treatment results from a field experiment in the Swiss Alps.”: This statement is an important introduction of your experiment, and as such, you should be more descriptive than "treatment results from a field experiment". What types of treatments specifically were involved in your field experiment, and were any of these treatments applied simultaneously to study interactive effects?

A > The sentences of the whole paragraph have been improved (“Here, we present …”). As demanded by the reviewer, this paragraph wraps up what we did in the experiment. We are of the clear opinion that more information at the end of an Introduction is inappropriate, as the full information is given in the Materials and Methods section.

83–“monoliths (ML)”: I do not feel that it is necessary to use an acronym for one word, and stating monolith regularly instead of ML will improve the clarity of your manuscript.

A > Agreed. We changed to “monolith” in the revised manuscript.

102-103: This sentence is rather unclear. What do you mean by standardizing harvests and the "zero-year" and "acclimation" distinctions? This aspect of your methods

deserves an elaboration.

A > Indeed, the distinction between ‘zero-year’ and ‘acclimation’ is obsolete. It derives from the chronology of establishing the experiment. The ‘standardizing’ harvests in these first two years served to homogenize the canopy of the monoliths, that were originally grazed and therefore had more heterogeneous canopies than mown grassland (new l. 182-183).

111-115: I find your naming convention, using the ’CS’ designations, to be needlessly confusing. These are simply sites along an elevational gradient, so why not refer to them either by their numeric elevation (i.e. 2360 m) or simply as Elevation 1 (lowest elevation), Elevation 2.... etc., instead of introducing a less intuitive naming system.

A > We chose the term ‘climate scenario’ (CS) to make clear that these sites are associated with a very complex treatment, containing a number of factors. Namely, the treatment includes changes in temperature, growing season length, and soil moisture. The index we chose (1-6) exactly follows your suggestion. We wish to keep this as it is.

116–“...6 CS, 6 MLs from each of the six sites of origin”: I find your naming convention, using the ’CS’ designations, to be needlessly confusing. These are simply sites along an elevational gradient, so why not refer to them either by their numeric elevation (i.e. 2360 m) or simply as Elevation 1, Elevation 2.... etc., instead of introducing a less intuitive naming system.

A > Please compare our response to the previous comment.

119–“...were filled with soil to prevent air flow”: Where did this soil come from? Bulk soil from each specific elevation/origin location?

A > The soil used originates from the respective scenario site, i.e. from the pit that was dug to accommodate the transplanted monoliths. This does not affect the individual turf monoliths’ soil properties, because the monoliths remained in their drained containers for the whole duration of the experiment, so that the monolith-soil was isolated both from neighboring monoliths. Because it seems that this detail is more confusing than helping, it has been deleted in the revised manuscript.

121–“cross-factorial design”: Full-factorial design? I’m unfamiliar with "cross-factorial" experimental designs.

A > This has been changed to ‘full factorial’ (new l. 216).

153: This sentence is rather unclear... Temperatures were summed across one day?

A > Unfortunately we can't find a reference to temperatures in l. 153. Our best guess is that this comment refers to l. 148 ff:
'The thermal energy was expressed as degree day values (DD0°C), resulting from hourly air temperature means above a threshold of 0 °C, added for one day, then divided by 24.'
Indeed, there is a plethora of 'degree days', tailored to suit many specific purposes and there is no single convention. The section has been improved to increase clarity (new l. 255-259).

> 154-156: This threshold seems particularly arbitrary, and I think that the use of a threshold in general is not necessary here. Why not simply present the mean growing season soil volumetric water content for each site/each season? This metric is much simpler and more intuitive for readers to understand and compare your results across the elevational gradient.

A > We considered using mean growing season soil volumetric water content and dismissed the idea. The reason is similar to the problems arising when using mean temperatures: Plants do not experience 'mean' water contents, when coping with environmental growth constraints. For example, when plants experience a wet month after a dry month, the mean soil moisture may suggest perfect growing conditions, when they were bad the whole time indeed.
We do not think that an increasing number for dry situations is less intuitive than a decreasing number for soil water content.

> 161-162: Why does the amount of precipitation added to each monolith vary between years?

A > The application of the irrigation treatment was determined by the occurrence of dry soil situations, which varied among years. Therefore, no changes have been made to text here.

> 168: Listing the chemical formula of ammonium nitrate is not necessary.

A > This is a detail, which we prefer to keep. Other Journals like 'nature geoscience' do it. The editor may decide on this, or the Journal's proof reading editor.

> 226: Is there some type of relationship between atmospheric N-deposition rates and elevation? Perhaps describe N-deposition rates across the entire gradient, not just at the middle and low points of your elevational gradient.

A > We only have data for the second highest site $CS2_{reference}$ (3.3 kg N ha^{-1} a^{-1}) and the lowest site CS6 (4.3 kg N ha^{-1} a^{-1}). This difference likely reflects the distance of the CS from the (agricultural) N-sources. CS6 (1680 m a.s.l.) is close to a village, $CS2_{reference}$ (2170 m a.s.l.) is further up the mountain.

236: What does non-continuous mean? Non-linear?

A > It is non-linear. The text has been improved following the reviewer's assumption (new l. 386).

239–"...only one third of the pre-harvest period was dry": It is definitely a misnomer to describe conditions of lower than 40% moisture content as "dry". In fact, in most alpine systems, 30% moisture content is considered ideal moisture conditions for optimal microbial activity (see Hawkes et al. 2017 PNAS for a relevant discussion related to respiration and soil moisture). I would highly suggest re-characterizing the way in which you describe soil moisture in this manuscript. Instead of creating a binomial in soil moisture conditions around an arbitrary 40% moisture content threshold, why not just describe average soil moisture across the growing season on a continuous scale, i.e. just state average growing season soil moisture for the pre-harvest period.

A > We agree that it would be advantageous to find a better term than 'dry' for sentences like this. We now use 'less soil moisture' and similar where appropriate.
As explained in the Materials and Methods section, the 40% threshold was neither chosen arbitrarily nor does the SWC < 40 % threshold imply plant growth limitation. Instead, it is an empirically developed contrast for differences in the soil moisture status between the CSs and between years. More time below the threshold simply means a 'drier period' in relative terms. This is also described in new l. 260-263. See also our response above to the same issue (response to 154-156)
We find the Hawkes et al. 2017 paper brilliantly describing the legacy of local climatic history on differential, local microbial adaptation. They find that microbial respiration is effectively locally specialized to soil moisture conditions. We could not discover references to plants, plant productivity, ideal moisture conditions or alpine sites.
With respect to the suggested use of average soil moisture values, please compare our comment on this issue above.

248-249: Because you describe soil moisture conditions in the previous section using percent dry days, we have no way of understanding how this transplantation effect on soil moisture conditions (described using VWC) might interact with your other treatments.

A > We think that there is a way of understanding the transplantation effect. In the section quoted, we state both the SWC for transplanted monoliths and the undisturbed grassland using 'average SWC'. It turned out that the difference was only 1% vol. SWC, meaning that there was next to no transplantation effect on SWC.

251: I would suggest that productivity is the more appropriate term, consistent with literature in this area of ecological research, to describe your response variable.

A > We prefer to keep ‘yield’. This type of grassland is maintained by mowing or grazing, and in an agronomic context, yield is fully understood and a correct term. Moreover, with our data we can only offer a crude proxy for ‘productivity’ (sensu net ecosystem productivity) because the harvestable part of the canopy is less than net ecosystem production. To avoid over-interpretation we prefer ‘yield’.

259: In order to show evidence to support this claim, I would like to see a figure and the related statistics that shows the relationship between the productivity effect size (productivity in transplanted monoliths - productivity in control monoliths that were reinstalled at the same site / standard deviation of productivity across all monoliths) regressed against the temperature difference from the monolith’s original climate and the transplanted climate. In other words, how much of the change in productivity is explained by change in temperature following transplantation?

A > The wording is changed in the revised text to avoid misunderstanding (new l. 409). We do not claim that temperature caused the significant differences in yield. Instead, we refer to the climate scenario (CS) because it is one of the strengths of our experiment that we simulate climate change in the mountains as complex climate scenarios, including simultaneous changes in thermal energy, growing period length, water availability and increased pollutant deposition. We are of the clear opinion that the metric suggested by the reviewer would be misleading.

In addition, as demanded by the reviewer, we have also broken down our analysis to individual, environmental parameters of CS, namely degree days (DD0°C) and < 40% SWC conditions (Fig. 2; fitted lines based on generalized additive models). The joint interpretation of both panels allows for a good assessment of possible drivers for yield changes over the 6 climate scenarios.

260-261: What does "intermediate warming" mean here? Describing this result as "monoliths that experienced X-Y degrees of warming by being transplanted to warmer climates at lower elevations relative to climate at their original location showed increases in productivity".

A > The text has been changed to ‘intermediate sites’ to avoid confusion. In general, ‘warming’ refers to the “altitude-related warming component’ of the CS, and the corresponding temperatures are given in Table 1. We think that – at this stage of the Results

491 section – the term ‘warming’ should be clear in the context of the study. Moreover, we have
492 added a formulation following the reviewer’s suggestions to improve clarity (new l. 410).
493
494 262-264: This sentence is confusing. 2016 was the year in which productivity, on
495 average, was highest, but this was only the case at two sites? These two statements
496 seem to contradict one another.
497 **A >** Here, we don’t say that it was only the case at two sites. In fact, all but one CS (CS5)
498 showed maximum yield in 2016 (see Tab. 4).
499 We replaced ‘both’ by ‘also’, to be more clear. We use the term ‘also’ to draw attention to a
500 counterintuitive situation: Despite transplantation into contrasting environments (cooler at
501 CS1 and substantially warmer at CS6), production of the maximum yield coincided with the
502 weather conditions of the same year.
503
504 298: The title of this section seems to not relate to the results described within the
505 section. You already stated that each elevational site is characterized by different
506 temperature and precipitation regimes in your methods and in previous sections of the
507 results. Should this section describe the relationship between productivity and climate
508 at each elevation?
509 **A >** Indeed, this section describes the relationship between biomass yield and those
510 environmental parameters (thermal energy and moisture) that we quantified for the individual
511 climate scenario (CS) sites. This is different from the approach that treats CS as categories
512 that integrate multiple climate change aspects. Accordingly, we have changed the title to
513 ‘3.2.5 Yield at climate scenario sites strongly relates to changes in thermal energy and soil
514 moisture
515
516 325-326: Are there examples of other papers whose conclusions about the use of
517 degree days instead of mean temperatures over the same time frame?
518 **A >** Particularly in environments with strong temperature contrasts (day/night,
519 summer/winter) like mountains or deserts, the use of DD does constitute a more valuable
520 metric for plant usable thermal energy. Similar to mean soil moisture values, mean
521 temperatures can be extremely misleading, because a sequence of hot and freezing
522 temperatures may well result in a comfortable average temperature that the plants have
523 never experienced.
524 Some examples for the use of DDs in the context of grassland research are
525 - Dukes et al., PLoS Biology 2005 (Jasper Ridge Experiment (CA))
526 - Fridley et al., Nature Climate Change 2016 (plant funct. strategies of 20 years UK grassland
527 warming)

- Wang et al., Ecology Letters 2020 (extremely dry Tibetan alpine grassland)
- Wilsey et al., Journal of Applied Ecology 2018 (42 US grassland sites)
- Zimmermann and Kienast, Journal of Vegetation Science 1999 (Swiss alpine grasslands)

333-341: This section would benefit from a description of why the authors suspect that warming beyond "intermediate warming" was not associated with the same boost in productivity that was associated with intermediate warming.

A > Agreed. This was only implicitly described in the original text. We have re-structured the paragraph and have added a sentence stating that ‘the comparatively low growth response suggests, that the water supply at CS6 has already reached a critically low level.’ (new l. 531-533).

337–“cockchafer (Melolontha melolonth) infestation: Please describe what this organism is and how it is relevant to variability in productivity.

A > The Cockchafer is a bug; its larvae feed on roots. When there are many, they may kill the vegetation. The Cockchafer is probably best known for its periodical mass flight-years. In these years, it is a major pest. We have added some information, but would prefer not to add more general biology because it will hinder the flow of reading.

347-349: Grammatical errors and diction in this sentence make it unclear.

A > Reformulated sentence to be clearer (new l. 541 ff).

358: I think this statement describes my point about eliminating your use of the "percent dry days" metric entirely... Your results, using this metric, prevent readers from relating the soil moisture conditions present in your experiment to soil moisture conditions elsewhere. Furthermore, describing soil moisture conditions less than 40% as "dry" is a misnomer.

A > We admit that between-experiment comparisons of soil moisture conditions, or rather the water availability for plants, is close to impossible. The reasons are that
A) different plants have different capacities to exploit the moisture resource. That means that a species from one experiment thrives well at the same SWC when a species from another experiment does not.
B) different soils have different water potentials (osmotic plus matrix potential). As a result, soils with the same vol. % SWC may have totally different water availabilities from a plant perspective.

564 Consequently, we chose to generate a wide range of within-experiment soil moisture
565 conditions for comparison, rather than refer to literature values. Moreover, we believe that
566 quantifying environmental conditions by describing them as more or less dry gives the reader
567 a good idea of which situation was more beneficial and which was less so.
568 As stated above (reviewers comment l. 239) we agree that '*dry* days' is not a perfect choice.
569 Please see our answer there to the use of this term.
570
571 380: What caused increased evapotranspiration at CS5? Is it possible that too much
572 rainfall, either ambient or added as part of your irrigation treatment, could cause
573 leaching of important soil nutrients, with higher VWC leading to lower productivity? This
574 might be especially relevant in monoliths that received both an irrigation and
575 fertilization treatment.
576 **A >** We strongly assume that higher temperatures, in those climate scenario sites (CS3,
577 CS4, CS5 and CS6) downslope from our reference site $CS2_{reference}$, caused higher
578 evapotranspiration.
579 We have no reason to assume that there was too much rain. The nearby federal meteorology
580 station recorded 662 mm/year during the experiment, while the 1981-2010 mean is 706
581 mm/year. At the AlpGrass experimental site, geography implies a rather continental climate,
582 insofar as inner-alpine valleys like the Engadin are generally quite dry. Please also compare
583 Tab. 2.
584 Our irrigation treatment only added 12-21% of the seasonal rainfall, and the nitrogen
585 deposition treatment was equivalent to 20 mm precipitation per year for all monoliths. This is
586 not a likely scenario for nutrient leaching.
587
588 399-402: These are the only lines of this section of your discussion that reference your
589 results directly. These sentences should be moved up in this section, and you should
590 eliminate the references to other experiments with results that contradict what your
591 experiment found, as this section is very unclear as curssrently written. Which of these
592 citations and theories help explain your results? Remove the rest.
593 **A >** The respective sentence ('…we found no significant overall effect of N-deposition on
594 yield after five years …') has been moved up to the start of this paragraph. We agree that
595 this paragraph can be more concisely written. It has been boiled down to the most essential
596 statements.
597
598 426–"This implies that subalpine grassland productivity has likely not increased during
599 the past century warming": This statement is in no way supported by your results.

A > We found that those monoliths that were subjected to a cooling treatment (at CS1), such that they experienced the temperature conditions of the 1920s, did not show a reduced growth compared to the climate scenario at $CS2_{reference}$ with todays' temperatures. Given this data, it can reasonably be assumed that the last 100 years of warming did not affect plant growth yet.