Interactive comment on "Asymmetric Responses of Primary Productivity to Altered Precipitation Simulated by Ecosystem Models across Three Long-term Grassland Sites" *by* Donghai Wu et al.

Responses to the Comments of the Anonymous Referee #2

Please note that our responses are shown in black while the comments of the reviewers are in blue.

The manuscript "Asymmetric Responses of Primary Productivity to Altered Precipitation Simulated by Ecosystem Models across Three Long- term Grassland Sites" presents a smart and well-thought out study to evaluate the performance of a large range of ecosystem models in their abilities to represent grassland productivity under changing climatic conditions. This study provides much needed insights in how ecosystem models perform when compared to field observations and highlights research needs to make such models more useful for climate change studies.

The abstract and introduction section is very nicely written and tightly structured. Unfortunately, I found that the result and discussion sections didn't follow this nice and logical structure.

Responses: We greatly appreciate the reviewer's pertinent feedback and valuable comments on our study. We thank you for your time and effort in helping us to improve this paper. In our revised version, we conducted a lot of work in responses to all these profound suggestions. In particular, we have (1) reorganized the main results in the abstract part; (2) reorganized the four specific objectives in the introduction part; (3) clarified the metrics of the response of productivity to precipitation changes following the four specific objectives in method part; (4) reorganized and strengthened the result part following the four specific objectives; (5) reorganized and strengthened the discussion part following the four specific objectives; and (6) rewrote the conclusion part to be more focused on our main results. We now consistently present first the comparison between spatial slopes and temporal slopes; which is followed by the asymmetric responses of productivities to precipitation under normal and extreme conditions using two indices (asymmetry index from inter-annual productivity and precipitation, and sensitivity of productivity to altered rainfall conditions). The curvilinear responses of productivities to altered precipitation across the three sites by each model are presented last. The first three specific objectives follow the structure used in Knapp et al. (2017), who have established a conceptual model for the precipitation-productivity relationships.

In addition, we have redrawn the figures to make them clearer, we have added uncertainty estimates for observation-based asymmetry index, and we have also added uncertainty estimates by models. Detailed responses are as follows under each of your comments.

Knapp, A. K., Ciais, P., and Smith, M. D.: Reconciling inconsistencies in precipitation-productivity relationships: implications

for climate change, New Phytologist, 214, 41-47, doi:10.1111/nph.14381, 2017.

Most of the results do not fully account/present uncertainty estimates. Some do, but it is often insufficiently explained what measure of uncertainty/variation is presented. This makes it at time difficult to appropriately evaluate the relevance of patterns found in the results.

Responses: In our revised results and figures, we have added uncertainty estimates by models, which presenting the model uncertainty ranges using interquartile spread of the sensitivities between individual simulations (10^{th} and 90^{th} percentiles).

In addition, we have added the uncertainty estimates of observation-based asymmetry index. The method has been clarified in our method part (page 6|line 24 to page 7|line 18). We have also added the uncertainty estimates of observation-based temporal slopes using a bootstrap sampling method with 1000 replicates of the ANPP and precipitation time series. We can obtain 1000 temporal slopes for each replicate and present the observed uncertainty ranges using interquartile spread of the temporal slopes between individual replicates (10^{th} and 90^{th} percentiles).

I miss an explicit discussion of the potential discrepancies in spatial scale between observations and model simulations. These can be particularly relevant for often fine-scale heterogeneity in soil moisture dynamics. I also miss a discussion on the caveats of the specific approach that was used for precipitation manipulations (fixed percentage increase/decrease for each rainfall event). It is not clear that this is what is happening under climate change; and precipitation event size distribution has large impacts on soil moisture dynamics (e.g., Lauenroth, W.K. & Bradford, J.B. (2012) Ecohydrology of dry regions of the United States: water balance consequences of small precipitation events. Ecohydrology, 5, 46–53.

Responses: We thank the referee for this insightful remark. In our revised version, we have added discussion on these issues in the second paragraph of section 4.3 related to the uncertainties, knowledge gaps and suggestions of future work. The second paragraph of section 4.3 have been revised as follows:

Although the carbon-water interactions in current models have been improved during the last decades, there still exist large gaps for accurately diagnosing the errors in the representation of key processes and parameterizations. Suggestions that should be considered in future studies aimed at model-data interaction include: (1) Simulation of SWC in the soil layer(s) for which experimental data are available. This can provide insights into the bias of modeled sensitivities to precipitation and check explicitly the sensitivity of vegetation productivity to change in SWC; (2) more experiments are needed that assess also BNPP in order to evaluate the corresponding processes in models (Luo et al., 2017; Wilcox et al., 2017); (3) there still exist large gaps between changes of precipitation occurrence and intensity in reality and how we simulated them in the current work, i.e., the altered rainfall forcing datasets were constructed by decreasing/increasing the amount of precipitation in each precipitation event by a fixed percentage during the time-span of productivity observations at each site and not by modifying precipitation structure or reproducing the real treatment. Further studies need to consider better different scenarios of precipitation occurrence and intensity under climate change (Lauenroth and Bradford, 2012), which will likely help to better understand the responses of productivities to altered precipitation in the next decades. In addition, modelers will need to simulate the control experiments corresponding to the real local precipitation manipulations applied by field scientists, e.g., considering the observed time series of modified precipitation and vegetation composition, root profiles, nutrient cycling, phenology and carbon allocation as close as possible to local conditions. This should be a priority for future model-experiment interaction studies.

Specific comments:

Introduction

- page 3, lines 13-15: rephrase to make the assumption explicit that "adaptation of plant communities over long time scales" is adaptation to typical "water received from rainfall for growth" – and not just any amount of water

Responses: The expression has been revised to: "a 'vegetation constraint' reflecting the adaptation of plant communities over long time scales in such a way that grasslands make the best use of the typical water received from rainfall for growth."

- page 3, lines 13-17: placing all citations at the end can be interpreted that all these citations only support point 2 and that there is no citation to support point 1

Responses: The citations have been revised.

- page 3, lines 17-20: The argument why temporal relationships are more informative for climate change impacts studies than spatial ones is not clear to me. It seems that effects of climate change on ANPP have not only a temporal trend (as stated here), but also include changes in species (and their adaptations) due to migration/extinction when tracking climate – thus spatial patterns may also be relevant if chosen carefully to reflect projected climate differences.

Responses:

Here, we preferred the temporal models than spatial models for projecting the effect of climate change on grassland productivity because substantial changes in plant communities (turnover of dominant life-forms) and corresponding alterations in soil biogeochemistry only occur over long time scales (decades to centuries). However, even over long time scales, the future climate and interactions with other global change drivers are expected to lead to communities that do not match current spatially observed climate-vegetation patterns. Thus, at least for near to mid-term forecasts of climate change effects on productivities, temporal models are more representative than spatial models (Knapp et al., 2017). In our revised version, the sentence has been revised to: "For projecting the effect of climate change on grassland productivity in near to mid-term (coming decades), inter-annual relationships are arguably more informative than spatial relationships because spatial relationships reflect long-term adaptation of ecosystems, and because ANPP-P relationships from spatial gradients are confounded by the co-variation of gradients in other environmental variables (e.g. temperature and radiation) and soil properties (Estiarte et al., 2016; Knapp et al., 2017)."

Knapp, A. K., Ciais, P., and Smith, M. D.: Reconciling inconsistencies in precipitation-productivity relationships: implications for climate change, New Phytologist, 214, 41-47, doi:10.1111/nph.14381, 2017.

Materials and Methods

- page 4, lines 19-26: I am surprised by the selection of the three study sites: two are located in the USA and represent naturally occurring grassland ecosystems where fire is an integral part whereas STU is located in Europe and is a man-made habitat that otherwise would be forested. These stark differences should at least be mentioned and caveats discussed.

Responses: These three grasslands were selected because they lie along a mean annual precipitation (MAP) gradient, and have detailed meteorological data to force the models. While two are "natural" grasslands (KNZ and SGS) and one (STU) is not, global land surface models do not typically differentiate regarding the origin of ecosystem types and heavily managed grasslands and pastures represent a significant fraction of mesic grasslands globally. Semi-natural subalpine grasslands in the Alps were created several centuries ago, are very lightly managed and should be in equilibrium concerning soil physical conditions. It should be noted though that the grassland at STU is cut once a year and lightly fertilized every 2-4 years and in consequence differs in plant composition and soil fungi: bacteria ratio, which leads to different drought responses compared to abandoned grassland (Ingrisch et al., 2017; Karlowsky et al., 2018). Further, it is worth noting that the mesic grassland in the USA would also be forested if human-initiated prescribed fires were to be removed from the system (Briggs et al. 2005). Thus, these grassland sites lie along a continuum of dry natural grassland, mesic natural grassland maintained by human management, and anthropogenic moist grassland maintained by human management. We now use these descriptions when the sites are initially described.

Ingrisch, J., Karlowsky, S., Anadon-Rosell, A., Hasibeder, R., König, A., Augusti, A., Gleixner, G., and Bahn, M.: Land Use Alters the Drought Responses of Productivity and CO2 Fluxes in Mountain Grassland, Ecosystems, doi:10.1007/s10021-017-0178-0, 2017.

Karlowsky, S., Augusti, A., Ingrisch, J., Hasibeder, R., Lange, M., Lavorel, S., Bahn, M., Gleixner, G., and Wurzburger, N.: Land use in mountain grasslands alters drought response and recovery of carbon allocation and plant-microbial interactions, Journal of Ecology, 106, 1230-1243, doi:10.1111/1365-2745.12910, 2018.

Briggs, J. M., Knapp, A. K., Blair, J. M., Heisler, J. L., Hoch, G. A., Lett, M. S., and McCARRON, J. K.: An ecosystem in

transition: causes and consequences of the conversion of mesic grassland to shrubland, BioScience, 55, 243-254, doi:10.1641/0006-3568(2005)055[0243:AEITCA]2.0.CO;2, 2005.

- page 5, lines 24-25: Please provide some details on how the gap-filling was conducted and how much of the data were filled in – at least for precipitation. Various approaches can lead to considerable differences in precipitation values, e.g., seasonal biases in missing data.

Responses: Historical reconstructions of meteorological variables from gridded CRUNCEP data at 1/2 hourly time step, a reanalysis product that does not have gaps (Wei et al., 2014), were combined and bias-corrected with site observations to provide bias corrected historical forcing time series from 1901 to 2013 (CRUNCEP-BC). For example, the CRUNCEP precipitation data will be regressed against the SGS precipitation observations during 1986-2009 as $y=a*x+b$, where y is the SGS precipitation observations during 1986-2009; x is the CRUNCEP precipitation data at SGS during 1986-2009. Then, the long term forcing data in SGS can be produced by correcting CRUNCEP data (called CRUNCEP-BC, BC is for bias corrected) during 1901-2013 using the equation and fitted parameters of *a* and *b*.

- page 6, lines 20-25: * Why do you calculate the "median value of productivities in wet years with annual precipitation higher than the 90th percentile level" and don't simply take f(p95) = productivity value with annual precipitation at the 95th percentile? Aren't they the same? And equivalently for med(f(p10)) = f(p5)?

* It seems that AI simplifies to

** AI = (med(f(p90)) – mean(f)) / mean(f) – (mean(f) - med(f(p10)) / mean(f) # after inserting Rp and Rd and which simplifies to

** AI = (med(f(p90)) + med(f(p10))) / mean(f)

** AI = (f(p95) + f(p5)) / mean(f) # after inserting previous bullet point

* I don't understand why Rp and Rd are defined differently from each other and thus, AI is the sum instead of the difference between the 5%- and the 95%- quantiles. In most cases of somewhat symmetric distributions, f(p95) > mean(f) and mean(f) > f(p5) and thus AI > 0.

* Results presented for instance in Fig. 4 where AI < 0 and AI > 0 suggest that AI is calculated correctly, as I suggest here, but that the equation is incorrectly written.

* What is meant with "f is the inter-annual productivity" (line 22)? Isn't f simply equal to "annual productivity"?

Responses:

In this method, for example, we chose the $med(f_{p90})$, the median value of productivities in wet years with annual precipitation higher than the 90th percentile level, rather than directly using the value of productivity in wet year at the 95th percentile level because the latter method may produce an extreme value of productivity which is less representative of the relative productivity pulse in wet years for the particular site. In other words, we opted for this approach to avoid artefacts.

In order to characterize the asymmetry of productivity to precipitation, we define the asymmetry index (AI) from inter-annual productivity and precipitation data as follows:

$$AI = R_p - R_d \quad (1)$$

where R_p is the relative productivity pulse in wet years, and R_d is the relative productivity decline in dry years defined by:

$$R_p = (med(f_{p90}) - \bar{f})/\bar{f} \quad (2)$$

$$R_d = (\bar{f} - med(f_{p10}))/\bar{f} \quad (3)$$

Thus, the AI could be rewritten as follows:

$$AI = \frac{med(f_{p90})-\bar{f}}{\bar{f}} - \frac{\bar{f}-med(f_{p10})}{\bar{f}} = \frac{med(f_{p90})+med(f_{p10})-2*\bar{f}}{\bar{f}} \quad (4)$$

where f is the inter-annual productivity, being a function of environmental factors from models or observation; $\bar{f}$ is mean annual productivity in the period of measurements (Table 1); $med(f_{p90})$ is the median value of productivities in wet years with annual precipitation higher than the 90th percentile level; $med(f_{p10})$ is median value of productivities in all the dry years when annual precipitation is lower than the 10th percentile level.

In general, $R_p > 0$ indicates that the median value of productivities in wet years is higher than the mean annual productivity in the period of measurements; and $R_d > 0$ indicates that the median value of productivities in dry years is smaller than the mean annual productivity in the period of measurements. Therefore, if $AI > 0$, i.e., a positive asymmetry, means that there is a greater increase of productivity in wet years than decline in dry years; if $AI < 0$, i.e., a negative asymmetry, means that there is a greater decline of productivity in dry years than increase in wet years.

In our revised version, we have expanded the description of the asymmetry index to make this clear for readers.

Results

- The structure of the result subsections is unexpected. The research questions and methods are tightly structured around the estimation of parameters a and b of Eq. 1, of the asymmetry index AI, and of the sensitivity index S. The result section does not follow this layout. For instance, the first subsection 3.1 could be presented in terms of estimation (and uncertainty) of parameter a. Then, the subsection 3.2 contains really the results (with lacking uncertainty estimates) for parameter b – plus in its current form some results on CUE and ANPP/NPP which have not been motived/introduced so far (which is confusing). The topic of subsection 3.3 spatial/temporal relationships presents the results from the second objective (as listed in the last paragraph of the introduction section); however, the method section does not explain how the observations and simulated values were aggregated and compared to address this question.

Responses: In our revised version, we have reorganized and strengthened the results following the same logical structure as the introductory section. We now consistently present first the comparison between spatial slopes and temporal slopes; which is followed by the asymmetric responses of productivities to precipitation under normal and extreme conditions using two indices (asymmetry index from inter-annual productivity and precipitation, and sensitivity of productivity to altered rainfall conditions). The curvilinear responses of productivities to altered precipitation across the three sites by each model are presented last. The results have been improved in our revised version (page 8|line 8 to page 10|line 16).

- page 7, lines 23-24: I see little support in Fig. 1 for the claim of a "steeper curvature at STU despite saturation above ambient precipitation indicates a steeper decline of productivity for precipitation set below ambient for this site compared to KNZ and SGS (Fig. 1)" – the precipitation treatment at STU did not (or at most barely) cover the curved part of the fitted lines. In most cases, a horizontal line appears to have fitted the data better. The estimates of b remain imprecise for STU, but this uncertainty is unfortunately not quantified.

Responses: In our revised version, we have redrawn the figures to make it clear for readers (page 26). The fitted lines only covered the simulations under altered precipitation conditions. In addition, the expression "steeper curvature at STU despite saturation above ambient precipitation indicates a steeper decline of productivity for precipitation set below ambient for this site compared to KNZ and SGS" has been removed.

- page 8, line 11: How was the ensemble model result calculated? Is this the arithmetic mean, median, etc.?

- page 8, line 23: "median value of -0.12±0.11" – what does the error component "±0.11" represent? Is this the MAD?

Responses: The ensemble model result used the arithmetic mean value, and the model uncertainty ranges used interquartile spread of the asymmetry indices between individual simulations (10th and 90th percentiles) in our revised version.

- page 8, line 23: Why "proportionally" larger? I don't understand what does could mean, particularly, because Rd and Rp are both calculated relative to mean(f).

Responses: This sentence has been revised to: "Hence, for SGS simulated declines of GPP and NPP in dry years were larger than the increases in wet years."

- page 8, line 26: Why are the observed AI values presented without uncertainty estimate?

Responses: We agree. In our revised version, we have added the uncertainty estimates of observation-based asymmetry index. The method has been clarified in our method part (page 6|line 24 to page 7|line 18).

- Figure 4: The dots are too large relative to the figure; they are overlapping each other so much that it is really hard to see what is going on. For instance, the reported 0.1-"significance" with an unnamed test for STU seems dubious as the visible few dots huddle around 0.

Responses: In our revised version, we have redrawn the figure to make it clear for readers (page 24). Here, a nonparametric test called the Wilcoxon signed rank was used for the significance test.

- Figures S1 and S2: There are no error estimates for parameters a and b. At least add appropriate error bars to Figs S1-S2. I don't understand why a and b are presented against each other in a scatter plot. In my understanding, there is no expectation of a relationship between a and b. This is confusing.

Responses: In our revised figures, we have added error estimates for parameters *a* and *b*. Here, we just intended to present the parameters *a* and *b* in a two-dimensional space rather than to make a relationship between the two parameters.

- Figures S3 to S6: There is too much on these panels. It is no longer possible to identify responses of individual models.

Responses: In our revised version, we have redrawn the figure to make it clear for readers.

- Figure S4: Is it correct that the "P" responses represents Rp of Eq. 3 and that "D" represents Rd of Eq. 4. Make this clear and use consistent terminology throughout the manuscript.

Responses: In our revised version, we changed the "P" and "D" to "R_p" and "R_d" in the figure to make the expression consistently throughout the manuscript.

- Figures S5 and S6: How to the absolute SWC values compare between observed and simulated?

Responses: The two figures are used to discuss the curvilinear responses of productivities to altered precipitation by models rather than to compare with observations. In this first model-experiment interaction study, we did not have any observed SWC data under differently altered precipitation conditions. For comparing the observed and modeled SWC, we recommended that models should report SWC at the same depth of experiments and experimental data should be made available for better comparisons in following studies. This can provide insights into the bias of modeled sensitivities to precipitation and check explicitly the sensitivity of vegetation productivity to change in SWC.

- Figures S7 to S9: Error estimates are missing and would be crucial to compare between CN- and C-only models.

Responses: In our revised figures, we have added uncertainty estimates for C-N and C-only models, which presenting the model uncertainty ranges using interquartile spread of the sensitivities between individual simulations (10th and 90th percentiles).

Discussion

- page 9, lines 21-29: The first paragraph of the discussion reads like an introduction paragraph that identifies the knowledge gaps.

Responses: This paragraph has been removed in our revised version.

- page 11, lines 25-27: Not clear what is meant here with "arid and semi-arid grasses […] show relatively strong resistance". Does this refer to varying abilities of grass species to extract soil moisture held at increasingly higher tensions? If this were fixed values in models across sites, then the simulations models may produce too high sensitivities at the drier sites, particularly SGS.

Responses: Grassland root depth affects ecosystem resilience to environmental stress such as drought, and arid and semi-arid grasses that have extensive lateral roots or possibly deep roots show relatively strong resistance. We agree, this also means that arid and semi-arid grasses may show greater abilities to extract soil moisture under conditions of increasing water stress. However, most ecosystem models currently consider only two types of grasslands, C3 and C4 (Table S14) with fixed root profiles along with prescribed soil layers (Table S13). This is potentially unrealistic for semi-arid grass roots and can lead to models underestimating the accessible water and the resistance to drought. We agree, the models may also produce too high sensitivities to drought at the drier sites.

- page 12, lines 1-2: I am confused here: the text continues to discuss "asymmetric responses" and yet refers to Fig. 5 which presents results for the sensitivity index calculated as relative difference among different model runs. So, if this text does refer to result for S, then I don't understand the statement "responses for normal precipitation variability" either because S isn't calculated from "normal precipitation variability" (as are Rd and Rp), but from manipulated precipitation inputs. Thus, S seems to rather represent sensitivity to deviation from 'normal' precipitation.

Responses: In this work, we characterized the asymmetric responses of productivities to precipitation under normal and extreme conditions using two indices (asymmetry index from inter-annual productivity and precipitation, and sensitivity of productivity to altered rainfall conditions). The asymmetry index could present the asymmetric responses under normal conditions, and the sensitivity of productivity to altered rainfall conditions could suggest the asymmetric responses under both normal and extreme conditions (Knapp et al., 2017). In our revised version, the sentence has been revised to "The sensitivity of productivity to increased and decreased precipitation for simulations where mean precipitation was normally altered generally suggested negative asymmetric responses at dry (SGS) and mesic (KNZ) sites (Fig. 3c)."

Data availability - Why are (a relevant subset of) data not deposited in a repository such as Dryad or figshare?

Responses: The amount of data is very large (about 25 GB), and we will try to use an online repository for sharing the modeled outputs.