

Response to Reviews

Date: January 24, 2018

Dr. Buckley,

Thank you for the opportunity to revise our manuscript. We apologize for the delay of our resubmission. The holidays intervened and we caught some errors we wanted to fully correct before resubmitting.

All three reviewers agree our paper is interesting, but point out several areas where we could improve the analysis. We were impressed with the very constructive reviews, and thank the reviewers for taking such care when reading our paper. We hope our paper is now suitable for publication in *PeerJ*.

In our point-by-point response below, the original reviews are in black and our responses are in blue. Likewise, changes made in the manuscript text are in blue for clarity. In our responses, we refer to line numbers of the revised manuscript unless otherwise stated.

Manuscript number: 20739

Title: Consistent ecosystem functional response across precipitation extremes in a sagebrush steppe

Author(s): Tredennick, A.T.; Kleinhesselink, A.R.; Taylor, J.B.; Adler, P.B.

Editors Comments for the Author(s)

Please pay close attention to the comments from Reviewer 1 regarding the validity of the findings. Once you have responded to these comments in a point-by-point cover letter and using tracked changes in the manuscript document, I will likely send your manuscript out for re-review.

We provide point-by-point responses to reviewer comments below. We have taken care to address Reviewer 1's concerns (also echoed by Reviewer 2) by including the suggested sensitivity analysis (see new section in Methods, lines 261-273). Unfortunately, we could not implement the suggestion of Reviewer 1 to include a $VWC \times Treatment \times Year$ interaction in our model because our data limits the complexity of our statistical model. Within a year and treatment, each plot has the same value of volumetric water content. Therefore, we simply don't have the variation needed to fit a $VWC \times Treatment \times Year$ interaction, which would be necessary to incorporate the effect of time on the ecosystem functional response.

The sensitivity analysis allows us to examine how the magnitude of the treatment effect changes over time. We found that the response of ANPP to drought grew over time, while the response of ANPP to irrigation is static (new panel C in Figure 3). We think this analysis improves the connection of our results to the Hierarchical Response Framework.

In addition to changes in response to reviewer comments, we have added more detail in our description of how we estimated soil moisture based on direct observations (lines 149-186). While doing this, we discovered an error on our part: we use March-June volumetric water content as our covariate in each year, but we did not actually have soil moisture estimates for March

2012 (sensors were not installed until April 2012). Rather than dropping an important month in terms of water availability, we decided to use output from the SOILWAT soil moisture model as parameterized for our site (lines 176-186 in the revised manuscript, Figure 2B, and new Appendix 2). Importantly, our main results are unchanged: ecosystem functional responses is similar among treatments, though there is now weak evidence that the drought response is more steep than the control response.

Reviewer: 1

Basic reporting

The basic reporting is quite good. The manuscript is well-written and easy to follow. The statistics and pre-treatment analyses were, despite being complicated, easy to understand. No comments or concerns here.

Thanks!

Experimental design

The study was well executed. I particularly like the use of the Sala experimental design of automated drought+irrigation treatments. I think this is a powerful design that enables the authors to speak more thoroughly about the role of water limitation in this system (as opposed to a drought-only experiment). The author also did an extraordinary amount of work on the statistical analyses and to resolve issues (VWC measurements/ANPP calculations) in a quantitatively rigorous manner.

Thanks again!

Validity of the findings

Their results indicate that the sagebrush steppe is actually quite robust to interannual variability in precipitation. This comes as a surprise, as the literature suggests that arid and semi-arid ecosystems (into which the sagebrush steppe easily fits) are quite sensitive to changes in precipitation. Huxman et al. (2004) and Knapp et al. (2015) both showed that drought impacts semi-arid sites more intensely than mesic sites, and Wilcox et al. (2017, in GCB) showed that most systems are more sensitive to irrigation than drought, and this is especially true for semi-arid sites. So far as I can tell, this study doesn't fit neatly into either line of thought (since it was relatively insensitive to drought, and irrigation effects were actually weaker than drought - see Wilcox et al.). It would be beneficial for the authors to put their study into this larger context and perhaps discuss why they think their site doesn't conform to these broader patterns (i.e. shrub-dominated, etc.).

This is an excellent suggestion, and we have broadened the scope of our Discussion to place our findings in the context suggested by the reviewer. In particular, we discuss why our site might be different from the large scale patterns seen in other areas (lines 349-351 and lines 392-394).

I also have a couple of concerns that potentially influence the results that the authors need to address prior to publication.

1. I appreciate the authors effort in validating and calibrating their NDVI measurements to ANPP, but I am quite concerned about the accuracy of these calibrations. The R^2 is quite low for those calibration regressions (0.2 - 0.7), so I wonder how accurately biomass in each plot was represented by these ANPP measures. This is actually quite a large source of error, and I wonder if it isn't responsible for the lack of significance and weak responses observed

here. I have two thoughts: would it be better to use raw NDVI measures rather than ANPP conversions? could the authors imbed the ANPP-NDVI calibration into their regressions (see Kiona Ogle's papers on multiple datasets in Bayesian analyses)? This would allow full uncertainty around the estimated ANPP based on NDVI to propagate into the VWC analyses.

This is a thoughtful critique, and we thank the reviewer for pushing us to think deeper about our approach. We did initially consider embedding ANPP-NDVI calibration into the overall model, as suggested by the reviewer. But, in the end, we decided this would muddle the analysis and confuse the reader, which is why we kept it separate. In our revision, we have included a new appendix (Appendix 5) where we conduct our regression with NDVI rather than ANPP. We achieved the same results as when analyzing ANPP (lines 203-204), and keep ANPP in the main text because that variable connects more easily to previous research (and what readers might be expecting).

A beneficial consequence of this comment from the reviewer is that we thoroughly checked our code for converting NDVI to ANPP and found an error for one year. Now our lowest R^2 is 0.39, rather than 0.21 (see new Table A3-2). This is still low, which is why we conducted the comparison analysis described above.

2. I didn't see mention of the priors on β , but I might've missed it. See Lemoine et al. (2016) - Ecology for a discussion of this issue. (very minor issue, I doubt the priors make too much difference here).

Sorry this was not clear. We included the prior specification [$\beta \sim \text{Normal}(0,5)$] in Appendix 4 on line 58. We have now included this in the main text on lines 251-252.

3. My biggest concern is that the statistical tests don't seem to test the HRF because they treat year/time as a "nuisance" variable that adds noise to the observations. But the HRF is explicitly about time and how the response of ecosystem function changes over time. So incorporating time as an explicit part of the analysis would be necessary. The easiest way would be to regress ANPP against VWC and allow the slope to vary by year to see if the slopes follow the pattern dictated by HRF (or could use a hierarchical gaussian process/LOWESS model or other non-parametric model to allow for non-lienarity along the VWC gradient - drought - control - irrigation). Perhaps another approach would be to calculate a 'sensitivity' for each replicate (or treatment) within a Bayesian model, then look at how sensitivity to drought/irrigation changes through time (in the same model, i.e. plot-level sensitivity nested within year, with fixed effects to contrast sensitivity to drought/irrigation, *a la* Wilcox et al. 2017). That is an explicit test of HRF, where the sensitivities could show any number of temporal patterns outlined by Smith.

The reviewer is correct that our analysis is not an explicit test of the Hierarchical Response Framework because we did not explicitly model the effect of time on ecosystem functional response. Unfortunately, we are unable to do so given our data – in each year every plot shares the same average soil moisture for its particular treatment, making it difficult to fit year-specific slopes for each treatment level.

That said, the reviewer made an excellent suggestion of conducting a sensitivity analysis *a la* Wilcox et al. 2017, which we have included (see lines 261-273 in the Methods section; new results on lines 316-317; and new discussion items on lines 395-401). The sensitivity analysis allows us to examine how the effect of treatment changes over time, adding another dimension to our

analysis that better connects our findings to the Hierarchical Response Framework. We found some evidence that the ecosystem's sensitivity to drought increased over time. This increase was not large enough to significantly impact ecosystem functional response, and our community composition analysis does not indicate changes in the final years of our experiment. However, as we note in the final line of our discussion (lines 399-401), the increased sensitivity in the latter years of the experiment may portend a coming shift in ecosystem functional response.

Reviewer: 2

Basic reporting

- Well-presented manuscript. I appreciated the additional details provided in the appendices, particularly appendix 2.

Thanks. We're glad you found our model description helpful.

- A more general ecology audience may benefit if the introduction would start with an additional first paragraph to provide a framing of the research in a larger picture and to explain why the ANPP–soil moisture relationship is so important.

Thanks for the suggestion. We have added new introductory sentences (lines 44-46).

Experimental design

- I don't understand the rationale for why the analysis is presented in terms of ANPP responses whereas NDVI was measured. The ANPP values represent a linear transformation of the NDVI measurements, but with a mediocre to low R^2 (A1-2), e.g., R^2 for year 2015 was 0.21 – an information that is somewhat buried in the supplementary files. This less than ideal relationship is not discussed, e.g., as potential fourth source for not finding responses (lines 285-288). Why not analyzing NDVI–soil moisture relationships instead/in addition to ANPP–soil moisture relationships?

This comment echoes one of reviewer 1's concerns. As we mention above, we have conducted a model comparison where we show that results are consistent whether we use ANPP or NDVI (see new Appendix 5). In the main text we retain the analysis focused on ANPP because we think it connects better to previous research. Also, as mentioned above, there was an error in our regression for the year 2015, which has been fixed (see new Table A3-2).

Validity of the findings

- The interpretation/discussion is coherent with the results. Even though the study found no responses, I agree that this is a reasonable outcome for a shrub-dominated, semi-arid ecosystem.
- The introduction presents the idea that there are four basic types of how ANPP responds to increases/decreases in soil moisture and that the time it takes to express a response will differ among different communities (lines 87-90). Even though the discussion section suggests that species life history traits ('bet hedging', lines 310-319) are important to understand the results of no treatment responses, I feel that it would be as important to discuss how the lifespan of species in the community in relation to the experimental duration of the treatments promotes/hinders the observation of responses.

This is an interesting point, and one we failed to discuss. We have added sentences about this in our Discussion (lines 399-401) Conclusion section (lines 409-411).

- Several of the dominant species (e.g., threetip sagebrush and bluebunch wheatgrass) have a lifespan that is much longer than the 5-year experiment; while other species in the community have shorter reproductive cycles. It seems that even with a 5-year experiment the longer-lived species of the experimental community may not have had sufficient time to (fully) respond and make it possible to observe measurable changes in the community composition/response.

Yes, we agree that the rate of community turnover is likely related to the lifespans of constituent species. We have added text discussing the impact of lifespan on community turnover in our Conclusion (lines 409-411).

- I would find it interesting if the plant community analysis is additionally carried out separately for short-lived vs long-lived species (Fig. 4) – based on the expectations that short-lived species have more ‘opportunities’ to respond/turnover than long-lived species.

This is a great suggestion, and one we had not thought of. We conducted the analysis suggested by the reviewer and found that community composition was also stable when we restricted the analysis to annual plants (short-lived). In a few years we detected significant treatment effects on dispersion or mean of community composition. But upon further investigation, we found that these statistical results were detecting changes in rare species because the dominant annuals showed similar dynamics through time, regardless of treatment (Fig. A1-4). Likewise, the stable perennial community contributes most to ANPP in this systems. We briefly mention this analysis in our Results section on lines 322-326, and show the full analysis in Appendix 1.

Comments for the Author

- Fig. 3: Misspelling in ‘increasing kernal bandwidth’ – this should be ‘kernel’

Good catch. Fixed.

- Fig 4: Expand caption: explain that dots represents plots and dimensions represent species.

Good idea. Done.

Reviewer: 3

Basic reporting

English is high quality with very few typos, though a few locations in text could use rephrasing to increase clarity. Some restructuring of the introduction is suggested to provide a more logical and engaging presentation of the topic. Whenever I had a question about a topic the authors were referring to, or a method they were using, their citations were sufficient for me to conduct further investigation. More discussion of prior work into the role of dominant species in determining ecosystem function may be warranted.

These are all good suggestions. Reviewer 2 agreed that the paper would benefit from a broader introduction, which we have added (lines 44-46). We have also added a reference to the idea that dominant species play a large role in determining ecosystem functioning (line 376).

There are no significant departures from standard structure.

Overall figures are well designed and easy to read, but treatment colors don’t translate well into greyscale. More contrasting colors (in terms of light and dark) are suggested. Minor things about a couple of table and figure headings could be changed to improve clarity. See notes to authors.

We have increased the contrast between the colors so that it should be easier to distinguish colors if printed in black and white. The colors aren't perfect, but they are now colorblind friendly and we anticipate most people will read this paper online.

Data is easy to find and download.

We're glad this was easy to access.

Product is a self-contained research unit that doesn't leave the reader wondering where the rest of the results are.

Experimental design

Primary research in Environmental science.

The research goal is to investigate how the sagebrush steppe ecosystem may respond to long-term climate extremes outside of historic norms. Since most predictive work is currently done using long-term correlations between ecosystem function and climate, but the climate is changing this is a pressing concern. The inherent variability of ecological responses across ecosystems means that each ecosystem needs assessment.

Design of the experiment looks very much like it was completely specified ahead of time and largely adhered to throughout. Where adjustments to the initial design were made, (probably in the soil moisture modeling) the procedure by which this was done was completely laid out. Though this is one of the areas I thought could use some rephrasing to increase clarity. The modeling work is excellent and looks at many different aspects of the question.

We have revised our description of modeling/extrapolating the discrete soil moisture measurements (lines 149-175). Hopefully it is now easier to understand.

Generally, the methods are quite detailed, including the description of the statistical analysis so reproduction should be possible. There were a few sections I thought could use additional clarification and these have been addressed in the comments to the authors.

Thanks for reading our methods carefully and providing good suggestions for clarity.

Validity of the findings

Many of the findings of this paper were negative, i.e. no change in ecosystem function or composition was observed. The authors draw the appropriate conservative conclusions of no effect. They do however discuss the possible implications of the positive and negative lean in the data, which might make for interesting grounds for future research.

Sample size is not large, though neither is it tiny. It was analyzed in such a way that information could still be learned despite the smaller sample size. Appropriate controls were included, though I would like a bit more information on why they used the controls they did, see comments to authors.

Good question! We used the control plots we did because demographic data had been collected in these plots historically (e.g., the 1930's) and our other studies rely on comparing new demographic data to historic demographic data. We also use the demographic data to determine species composition. Given that this demographic data is time-consuming to collect, we decided to use the historical plots as controls rather than adding even more plots. We have added our reasoning on lines 119-120

Conclusions are clear and do not reach beyond the scope of the data.

Authors speculated at length regarding why negative results were found. Concluding that it was likely a property of the ecosystem and probably not an artifact of the way the experiment was carried out, though they were careful to state that this possibility could not be ruled out.

Comments for the Author

This is a good article that yields insight into how experimental work can allow us to test whether historic precipitation/biomass relationships will hold up to potential future changes in the precipitation regime. Overall, this is a very strong study and paper; it will make an excellent contribution to the scientific literature. However, there are some small issues that should be resolved first. There are some organizational issues in the Introduction; particularly, it seems the motivation for the study is buried deep within the text instead of in the first paragraph. Revising this section would make the paper a much more compelling read. Additionally, there were a number of places in the methods where I felt that additional clarity was needed so that others could replicate their study. Below I give line-by-line suggestions on these points and other minor editing suggestions. This reviewer feels that this paper should be accepted with minor revisions.

Thanks for these thoughtful comments. We address your specific points below.

General comments to authors

Lines 31-32: sentence beginning with “However, ANPP response...” seems fairly vague and weakly worded, consider revising to increase clarity.

We have revised this sentence to read, "However, all estimates for the effect of soil moisture on ANPP broadly overlapped zero, indicating the relationship is weak and uncertain regardless of treatment." (lines 31-32).

Introduction:

While the abstract opens with a strong justification tying the paper into shifting precipitation patterns as a result of climate change and the need to develop a way to predict how communities might change with long-term exposure, the introduction opens only with a statement saying that there is a relationship between soil moisture and ANPP. This seems incredibly weak and unconvincing. The authors should consider restructuring their introduction so that the justification for their study is at the beginning. They have several other paragraphs in the paper that could do nicely as an opening paragraph. The opening paragraph of the discussion Lines 262-269 would make for an excellent opening. Or, with a little retooling the second paragraph of the introduction (lines 53-64) would make for a better statement of research motivation than the current opening.

Fair point, and one that echoes Reviewer 2's critique of our introduction. We have added new opening sentences that we hope better set the stage (lines 44-46).

Line 46 “outside” is repeated twice

Good catch. Extra "outside" has been removed.

Lines 71-72. While this is an excellent and clear definition of ‘ecosystem functional response’, I think it might be prudent at some point to explain why the authors are using soil moisture as an explanatory variable for ANPP instead of [precipitation]. I agree that soil moisture is generally a better predictor of ANPP than direct precipitation measures, but most of the articles the authors

cite refer to ANPP-precipitation relationships, so taking a moment to explain why they are using soil moisture would be warranted.

Good point. We have added a sentence describing why we use soil moisture on lines 76-77.

Lines 88-90: another good motivating factor for why this research was done that could be used earlier in the introduction.

Thanks for the suggestion. We have alluded to this in our new opening paragraph (lines 44-46)

Final paragraph on introduction. I really like the way the authors are testing multiple hypothesis based on theory.

Thanks!

Methods:

110-113: Could the authors please explain why these long-term monitoring plots make better controls than just say a generic 1 m² plot within the exclosures next to the experimental plots. Certainly the long-term record on community shifts in these plots are likely interesting in their own right, but it doesn't seem like they add much to the current study.

From above, "We used the control plots we did because demographic data had been collected in these plots historically (e.g., the 1930's) and our other studies rely on comparing new demographic data to historic demographic data. We also use the demographic data to determine species composition. Given that this demographic data is time-consuming to collect, we decided to use the historical plots as controls rather than adding even more plots." We have added this justification in our Methods section on lines 119-120.

Lines 126-127: It does concern me that the authors didn't account for snow/winter precipitation on their irrigation plots. It appears as though a large percentage of the yearly precipitation in Dubois, ID falls over winter. Additionally, since their drought shelters were left up over winter they may actually have attracted drifting and captured more snow on the plots. To their credit, the authors did address this issue to some extent in the Discussion, lines 305-306.

It concerns us, too. But, unfortunately, it was not feasible for us to control snow/winter precipitation. As the reviewer notes, we do consider this a limitation of our study (lines 372-373).

Lines 132-135: Wasn't initially clear that these were manually collected data instead of another set of sensors. Maybe mention that they were collected by handheld probe at the beginning of the description instead of at the very end.

Done (line 141).

Lines 142-147: Description of standardization procedure is confusing. I had to read it three times before the procedure was clear to me. Consider rewording for clarity.

Upon re-reading, we completely agree! We have substantially rewritten this section for clarity and have added some mathematical notation to make our description precise (lines 149-175). Hopefully it is less confusing now.

Lines 148-158: I'm a bit confused on which plots this modeling was conducted for. Was it done only on those plots that had permanent sensors in place? Or when possible was it also extrapolated to include those plots that only had point in time measurements. Please clarify.

We have revised this section to make this clearer (lines 149-175).

Lines 168-174: I like the idea of using the two separate estimates of greenness and then using the one that predicts best every year. Very thorough.

Thank you.

Line 188: later, in figure 2, the authors refer to mean volumetric water content as “VWC” also. It would be best to use a different abbreviation for the two separate uses of volumetric water content: mean and cumulative.

This is a good catch. We actually only use *cumulative* volumetric water content, and we have made this consistent throughout the manuscript. We also updated the Figure 2 caption to make this more clear.

Results:

Lines 241-244: Even though figures are referred to here, it might be nice to have some measure of the effect size in text, perhaps percentages, just to give a numerical estimation of differences, or lack there of, between plots in ANPP and soil moisture.

Good idea. We have added the effect sizes (lines 298-302).

Lines 257-260: the stability of the community over the study was impressive particularly given a method sensitive to rare species was used. It would be nice to see a bit more description of what the community looked like. How much of the community was the dominants, vs. rarer species?

We have added a new appendix (Appendix 1) that provides more details on the dominance and rarity of the communities across and through time. In brief, rank abundance curves are very similar for all treatments and through time, with the same two perennial species (*Artemisia tripartita* and *Balsamorhiza sagittata*) dominant in all treatments, accounting for about 25% of total cover. Similarly, all treatments had the same dominant annual species, whose dynamics did not vary by treatment. These ancillary results add weight to our conclusion that the ecosystem responses is driven by dominant species, which are similar among treatments. We mention these results in the manuscript on lines 322-326.

Discussion:

Lines 279-281: I like that the authors were very conservative in their eventual conclusions despite their discussion of the potential meanings of the parameter estimates when these simply leaned positively or negatively and all overlapped zero.

Thank you.

Lines 289-299: strong discussion and support for the sizes of the treatment applied. Good work.

Thanks for the kind words.

Lines 320-325: seems like authors are indicating that ecosystem response was driven by the response of the dominant species. Might be good to bring in some other work that has found similar results and discuss it a bit, for example: “Drivers of Variation in Aboveground Net Primary Productivity and Plant Community Composition Differ Across a Broad Precipitation Gradient” (La Pierre et al., 2016).

Thanks for bringing this paper to our attention. We have included a new sentence that references this citation on lines 367-368 in the Discussion.

Figures and tables:

Table 1: These are the beta coefficients from the model correct? Might be good to elaborate a bit more in the table description.

Yes, these are the beta coefficients. We have added detail to the table description.

Figure 2: Here is the use of VWC as “mean soil volumetric water content” I referred to earlier. Like I said it would probably be best to have different abbreviations for different metrics of soil moisture. Also interesting, though not the point of this paper, after three consecutive dry years with progressively decreasing ANPP it looks like the ecosystem was primed for recovery as soon as an average precipitation year hit in 2015. No evidence of drought legacy.

Sorry about this mistake. We use *cumulative* March-June volumetric water content for the statistical model, but in this figure we plot the estimated daily values as a time series. We have updated the figure axis label and the caption to clarify.

Figures general: differences between colors used to indicate treatments are not at all distinguishable if viewed in greyscale. Authors should consider using a color scheme with more contrast.

We have increased the contrast between colors so that they are easier to distinguish if printed in black and white. They are not perfect, but they are from a color-blind friendly palette and we anticipate most readers of *PeerJ* will read the paper online.