Reports

Ecology, 99(10), 2018, pp. 2145–2151

A reality check for climate change experiments: Do they reflect the real world?

Alan K. Knapp,[1,4] Charles J.W. Carroll,[1] Robert J. Griffin-Nolan,[1] Ingrid J. Slette,[1] Francis A. Chaves,[1] Lauren E. Baur,[2] Andrew J. Felton,[1] Jesse E. Gray,[1] Ava M. Hoffman,[1] Nathan P. Lemoine,[1] Wei Mao,[3] Alison K. Post,[1] and Melinda D. Smith[1]

[1]*Graduate Degree Program in Ecology and Department of Biology, Colorado State University, Fort Collins, Colorado 80523 USA*
[2]*Department of Biology, University of New Mexico, Albuquerque, New Mexico 87131 USA*
[3]*Northwest Institute of Eco-Environment and Resource, Chinese Academy of Sciences, Lanzhou 730000, China*

Abstract. Experiments are widely used in ecology, particularly for assessing global change impacts on ecosystem function. However, results from experiments often are inconsistent with observations made under natural conditions, suggesting the need for rigorous comparisons of experimental and observational studies. We conducted such a "reality check" for a grassland ecosystem by compiling results from nine independently conducted climate change experiments. Each experiment manipulated growing season precipitation (GSP) and measured responses in aboveground net primary production (ANPP). We compared results from experiments with long-term (33-yr) annual precipitation and ANPP records to ask if collectively (n = 44 experiment-years) experiments yielded estimates of ANPP, rain-use efficiency (RUE, grams per square meter ANPP per mm precipitation), and the relationship between GSP and ANPP comparable to observations. We found that mean ANPP and RUE from experiments did not deviate from observations. Experiments and observational data also yielded similar functional relationships between ANPP and GSP, but only within the range of historically observed GSP. Fewer experiments imposed extreme levels of GSP (outside the observed 33-yr record), but when these were included, they altered the GSP–ANPP relationship. This result underscores the need for more experiments imposing extreme precipitation levels to resolve how forecast changes in climate regimes will affect ecosystem function in the future.

Key words: aboveground net primary production; climate change; extreme climate; field experiments; grasslands; long-term data; precipitation.

Introduction

In his 1960 address to the Ecological Society of America (ESA) membership, ESA President Thomas Park called for an "acceleration" of the use of experiments to enhance "research progress and the validity of our interpretations" based primarily on observational studies (Park 1961). In the 50+ years since Park's address, experiments, particularly those conducted under field conditions, have become an integral and perhaps even dominant approach for understanding ecological pattern and process (Tilman 1989, Roush 1995, Ives et al. 1996, Knapp et al. 2012*b*). This is particularly true in global change ecology, where key drivers of ecological processes are expected to be altered in ways that differ from both historical and present-day environmental conditions (Williams and Jackson 2007, Thompson et al. 2013). For climate change research, observations made across natural climatic gradients or through time (e.g., Elmendorf et al. 2015, Mayor et al. 2017) can provide valuable insight. But the primary approaches ecologists use to understand how ecological processes may change in the future include manipulating climate drivers (e.g., temperature and precipitation) at individual sites, across natural gradients (e.g., Wu et al. 2011*b*, 2012), and via simulation modeling (Luo et al. 2011*a*, *b*).

Despite differences in the types of inference drawn from each approach, experimental and observational studies are viewed generally as complementary (Silvertown et al. 2010). Indeed, the value of interpreting experimental results within the context of observational data is clear in many contemporary global change studies (e.g., Hoover et al. 2014, Copeland et al. 2016, Liu et al. 2018). However, despite evidence that consistent inferences can be drawn from experimental, monitoring, and gradient approaches (Elmendorf et al. 2015), such complementarity has been challenged by experimental results directly contradicting observations (e.g., Sandel et al. 2010, Blume-Werry et al. 2016, Wardle 2016, Yuan et al. 2017, Barner et al. 2018). Not surprisingly, inconsistencies have been attributed both to experiments strongly underestimating observed climate effects (Wolkovich et al. 2012), as well as monitoring and gradient approaches overestimating responses to climate change (Metz and Tielbörger 2016).

Manuscript received 23 May 2018; accepted 5 July 2018. Corresponding Editor: Nathan J. Sanders.
[4] E-mail: aknapp@colostate.edu

There are many reasons that results from climate change experiments may not reflect observations made under natural conditions (Metz and Tielbörger 2016). For example, ecological responses to alterations in precipitation, one of the most commonly studied forecast climate changes (Wu et al. 2011*a*), have been evaluated with both observational and experimental approaches. Observational studies assess how ecological responses vary as precipitation amount varies over time and across geographic gradients (Knapp et al. 2017*b*); however, these analyses fail to account for a wide range of covarying factors (Estiarte et al. 2016, Metz and Tielbörger 2016), which can make isolating precipitation effects difficult. Experiments alleviate this concern, but are subject to a number of potential artifacts (Leuzinger et al. 2011, Beier et al. 2012, Wolkovich et al. 2012, De Boeck et al. 2015, Elmendorf et al. 2015, Hoover et al. 2018). Experiments, for example, often manipulate key ecological drivers in a simplistic manner, failing to capture the complexity of the real world (Kreyling and Beier 2013). Experiments also suffer from issues of scale (Sandel and Smith 2009) and tend to focus on short-term mechanisms that may not drive longer-term outcomes (Strauss et al. 2008, Saccone and Virtanen 2016). Despite these and other well-known limitations (Tilman 1989), we rely on climate change experiments to provide insight about the future, and in many cases to parameterize and improve models integral for forecasting (Luo et al. 2011*a*, *b*).

This raises an important question: How well do climate change experiments reflect reality? This query cannot be addressed by comparing the outcome of a single experiment to an observational data set. For example, we know from long-term observations that aboveground net primary production (ANPP) is strongly influenced by precipitation inputs, particularly in grasslands (Sala et al. 2012, Knapp et al. 2017*b*). However, in all such long-term data sets, there are many pairs of years in which ANPP does not conform to the expected response to precipitation (e.g., a drier year is more productive than a wetter year). Similarly, results from an individual experiment that includes water addition or removal may or may not support precipitation amount as a primary control on ANPP for any number of reasons (Cherwin and Knapp 2012, Hoover et al. 2018). Rather, to assess rigorously how well climate manipulation experiments reflect reality, results from multiple independently conducted experiments should be compared to many years of observations. Here, for the first time, we present such a comparison, one that takes advantage of multiple precipitation manipulation experiments conducted in a water-limited native grassland, the tallgrass prairie of North America (Knapp et al. 2001). We challenged these experiments to replicate reality (i.e., observational data) in two ways. First, we asked if the values of ANPP estimated from this collection of experiments differed from ANPP derived from long-term observations. In other words, if we had to rely on experiments to estimate ANPP, could they substitute for long-term observations? A second more substantial challenge involved asking if the functional relationship between ANPP and precipitation estimated from experiments could match long-term observations. In other words, we compared how alterations in precipitation amount altered ANPP based on experimental vs. observational data sets. Our analysis was based on research conducted at the Konza Prairie in northeast Kansas, USA, a Long-term Ecological Research (LTER) site with a multidecadal record of monitoring ANPP and precipitation. Over this same time period, a wide range of precipitation manipulation experiments have been conducted at Konza Prairie as well. Collectively, the experiments we compiled provide >40 estimates of ANPP responses to alterations in precipitation for comparison with long-term (14–33 yr) observations of ANPP and precipitation.

Methods

We compiled ANPP and precipitation data from control and treatments plots from nine experiments conducted at Konza Prairie (USA) over a 35-yr period (Table 1). Results from most of these experiments have been published, and data were extracted from these publications or the authors were contacted for the raw data. Data were also included from two unpublished experiments (Table 1). For some experiments, only mean values of ANPP and precipitation were available. Thus to be consistent, we only analyzed mean ANPP responses to precipitation for all experiments. Each experiment manipulated inputs of water by differing amounts, with some experiments imposing a single treatment level and others including multiple levels of precipitation. Moreover, these experiments used a variety of approaches to alter precipitation (e.g., passive and active; Knapp et al.

Table 1. Sources and years of data collection for aboveground net primary production (ANPP) and precipitation data from precipitation manipulation experiments conducted on Konza Prairie, Kansas, USA as well as observational data from long-term monitoring at multiple sites across the study site.

Source	Time period	Data
Experimental data		
Knapp (1984)	1982–1983	published data
Knapp et al. (2001)	1991–1992	Konza data catalog (ID: WAT01)
Fay et al. (2003)	1998–1999	www.konza.ksu.edu/ramps/data.html
VanderWeide et al. (2014)	2009–2010	published data
Hoover et al. (2014)	2010–2011	contributed by Author
Wilcox et al. (2015)	2011–2012	contributed by Author
Denton et al. (2017)	2013	contributed by Author
EDGE (*A.J. Felton, unpublished data*)	2013–2014	edge.biology.colostate.edu
Felton et al. (*A.J. Felton, unpublished data*)	2015–2016	contributed by Author
Observational data		
Konza LTER (upland)	1984–2015	Konza data catalog (ID: PAB01)
Konza LTER (lowland)	1984–2015	Konza data catalog (ID: PAB01)
RaMPs control plots	1998–2016	www.konza.ksu.edu/ramps/data.html
Briggs and Knapp (1995)	1984–1997	published data

Notes: Note that some experiments included multiple levels of precipitation treatments (increases and decreases) so that there were 42 estimates of ANPP responses to manipulated precipitation. Similarly, because of overlap in the years ANPP was measured at Konza Prairie, there were 95 observations.

2017*a*), and precipitation was altered for different portions of the growing season and often for multiple years. However, when experiments altered precipitation for many years consecutively, we only included responses from the initial two years. This was done because although it is common to encounter two consecutive wet or dry years in the observational data set ("reality"), three consecutive wet or dry years were uncommon. In addition, this minimized alterations in ecosystem function caused by the cumulative effects of increased or decreased precipitation (Smith et al. 2009, Smith 2011, Knapp et al. 2012*a*, Wilcox et al. 2016). Finally, no experiments manipulated water outside the growing season, and >90% were conducted in sites burned in the spring prior to the experiment. Because fire alters the water relations of grasslands and is essential for the maintenance of tallgrass prairie (Knapp et al. 1998), we only included experiments conducted in annually or frequently burned sites.

Long-term observational ANPP and precipitation data from annually burned grassland were combined from four data sets that encompassed a variety of topographic positions and soil types at Konza Prairie (Table 1). This was desirable because experiments were conducted in both uplands and lowlands, and from locations arrayed across the 3,487-ha site where soil types vary (Knapp et al. 1998). This helped insure that observational data were consistent with experiments with respect to capturing within-site variability. For all experiments and the long-term observational data, ANPP was estimated by harvesting aboveground biomass in late August–early September. In burned grassland sites, this represents peak aboveground biomass produced that year (Briggs and Knapp 2001). Finally, because experiments only manipulated precipitation during the growing season, we used the same period (1 April–31 August) for calculating growing season precipitation (GSP) for both experimental and observational data.

As noted above, we asked if mean ANPP for Konza Prairie based on long-term observational data differed from mean ANPP determined from experimental plots. This might be expected given that experiments were not conducted every year and thus experiments sampled only a subset of the years (15 yr) included in the observational data (33 yr, Table 1). Mean ANPP from experiments was calculated from control plots as well as by combining control and treatment plots. We also compared mean GSP in observational vs. experimental data sets. Using both GSP and ANPP data, we calculated rain-use efficiency (RUE) as the ratio of ANPP to precipitation ($g \cdot m^{-2} \cdot mm^{-1}$) for both data sets. Statistical significance was assessed with ANOVA ($\alpha = 0.05$, R version 3.4.3) (R Core Team 2018).

Assessing ANPP–GSP relationships from observational and experimental data that differed in the absolute magnitude of ANPP (due to differences in topographic position or soil type; Fig. 1a) was more challenging. Here, we calculated proportional responses in ANPP to precipitation to allow us to compare directly the slopes of the ANPP–GSP relationships. For observational data, we calculated mean ANPP and GSP across all years and sites and expressed interannual deviations from these means as a proportional change. For example, mean GSP was 529 mm for the observational data and a year with 610 mm GSP was represented as a 15% increase in GSP. The corresponding relative change in ANPP was determined similarly. For experimental data, we calculated the proportional increase or decrease in GSP and ANPP based on differences between treatment and control plots divided by control plot values. In this way, the sensitivity of ANPP to changes in precipitation could be compared (ANCOVA) between approaches using both linear and nonlinear regression analyses (R version 3.4.3).

Results

As reported previously, interannual variability in ANPP and precipitation (GSP) are strongly correlated in tallgrass prairie, and this was evident for each of the four long-term observational data sets we used (Fig. 1a). The slopes for these relationships were statistically indistinguishable (ranging from 0.47 to 0.55 grams per square meter ANPP per mm precipitation) indicating that sensitivity of ANPP to changes in GSP was similar throughout the site, regardless of the absolute level of productivity. We considered the range in GSP encountered over this 33-yr monitoring period (312–964 mm) as the nominal precipitation range for our analysis, and experiments (42 experiment-years total) provided data from 22 experiment-years that increased ($n = 14$) or decreased ($n = 8$) GSP within this range (Fig. 1b). A smaller number of experiment-years were available that increased ($n = 6$) or decreased ($n = 14$) GSP beyond this nominal range (hereafter referred to as extreme treatments). Bias toward extreme treatments that reduced rather than increased precipitation is not surprising given forecast increases in extreme drought globally (Smith 2011).

Based on observational data, mean ANPP and GSP for the study period was 511 g/m^2 and 529 mm, respectively (Fig. 2). This led to an overall site RUE of 0.99 ± 0.03 $g \cdot m^{-2} \cdot mm^{-1}$. We then estimated ANPP, GSP, and RUE from experimental data (1) by including only control plot data from experiments conducted within the nominal range of GSP, (2) by including nominal control and treatment plots (with GSP adjusted based on the treatment), and (3) by including all data (including extremes, Fig. 2). We calculated ANPP and RUE from other combinations of experimental data as well, but present these to capture the range of values estimated from just nominal control plots to all experimental data. In all cases, mean ANPP was statistically indistinguishable from the observational values (mean experimental ANPP = 493.91–502.32 g/m^2, $F = 0.187$, df = 3, 241, $P = 0.905$) as was GSP (mean experimental GSP = 487.97–512.90 mm, $F = 0.803$, df = 3, 241, $P = 0.493$). The same was true for all combinations of RUE from these experimental data sets (mean experimental RUE = 1.0–1.13 $g \cdot m^{-2} \cdot mm^{-1}$, $F = 2.29$, df = 3, 241, $P = 0.079$).

We next compared estimates of the sensitivity of ANPP to changes in GSP from both observational and experimental data sets. Because alterations in GSP in experiments exceeded those in the observational data, we estimated sensitivity (slope of the ANPP–GSP relationship) for experiments altering GSP within the nominal range first and then by also including extreme GSP alterations. Within the nominal GSP range, the estimated sensitivity of ANPP to alterations in GSP did not vary between the experimental and observational data sets (Fig. 3) and the relationship was linear. When results from experiments that altered GSP to

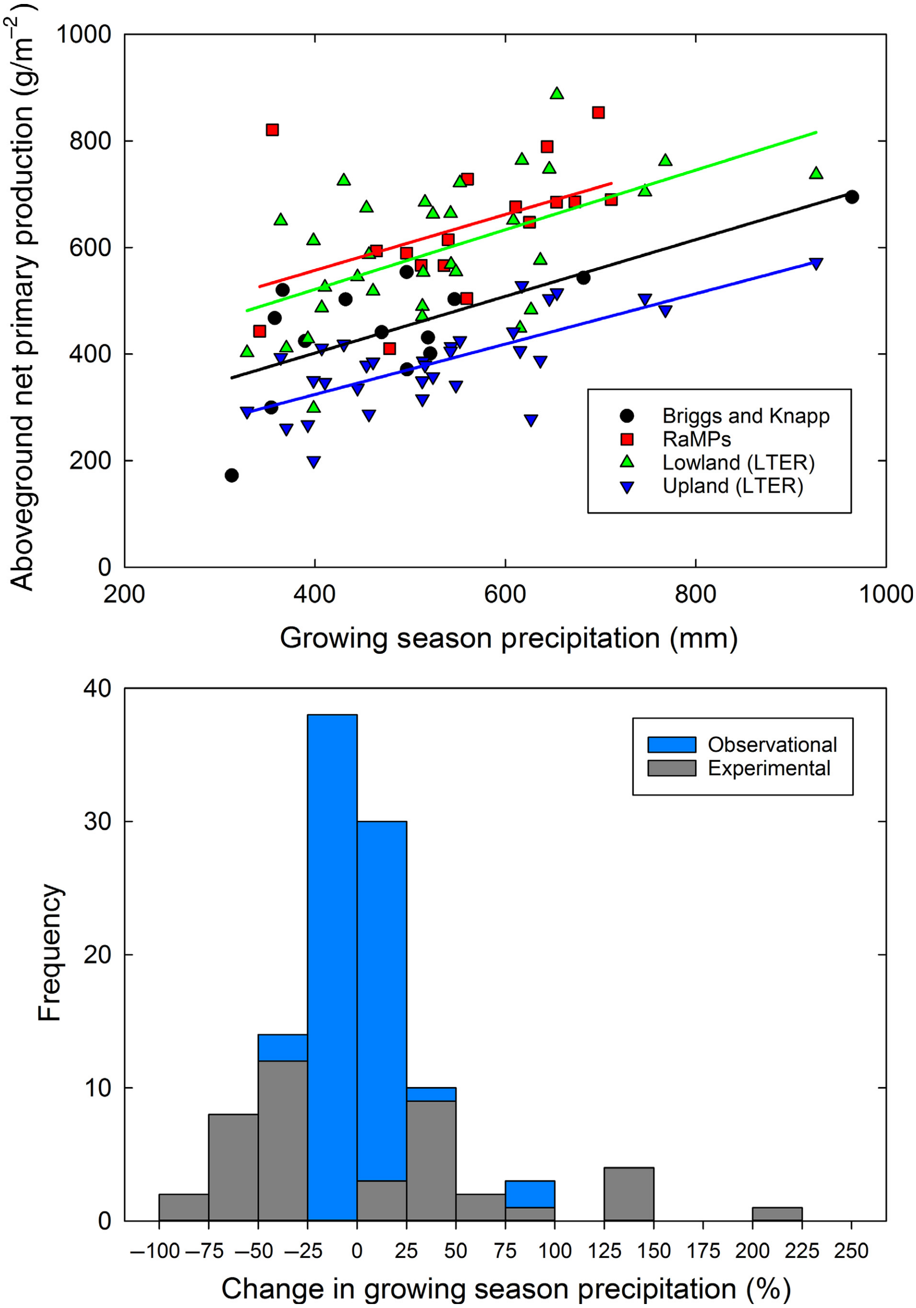

Fig. 1. (a) Relationship between growing season precipitation (GSP; 1 April–31 August) and aboveground net primary production (ANPP) for four long-term data sets at Konza Prairie, Kansas, USA. The two LTER data sets are from different locations in a single watershed whereas the Briggs and Knapp data set combines observations from multiple watersheds (see Table 1). RaMPs data represent long-term ANPP data from observations adjacent to the Rainfall Manipulation Plots (RaMPs) study (Knapp et al. 2001). Note that the ANPP–GSP relationship is identical for all data sets (slopes do not differ, $F = 0.082$, df = 3, 87, $P = 0.97$), despite differences in mean ANPP (intercepts differ). (b) Frequency histogram depicting the distribution of variation in GSP relative to the long-term mean for observational data and the distribution of differences between control and treatment GSP for all experiments combined (see Table 1).

extreme levels were included, the best-fit relationship was nonlinear, although the difference in AIC between linear and nonlinear relationships was small (367.45 and 365.15, respectively).

Discussion

This "reality check" for experiments, comparing results from long-term observations of ANPP and precipitation, our benchmark for reality, with a body of independently conducted experiments, provides three key insights. First, precipitation experiments in this grassland yielded estimates of ANPP and RUE, as well as the sensitivity of ANPP to changes in GSP, similar to those based on long-term observations (Fig. 2, 3). This is important given concerns that single-factor experiments do not realistically capture "real world" complexity, omit important interactions, and thus will over- or underestimate ecosystem responses (Leuzinger

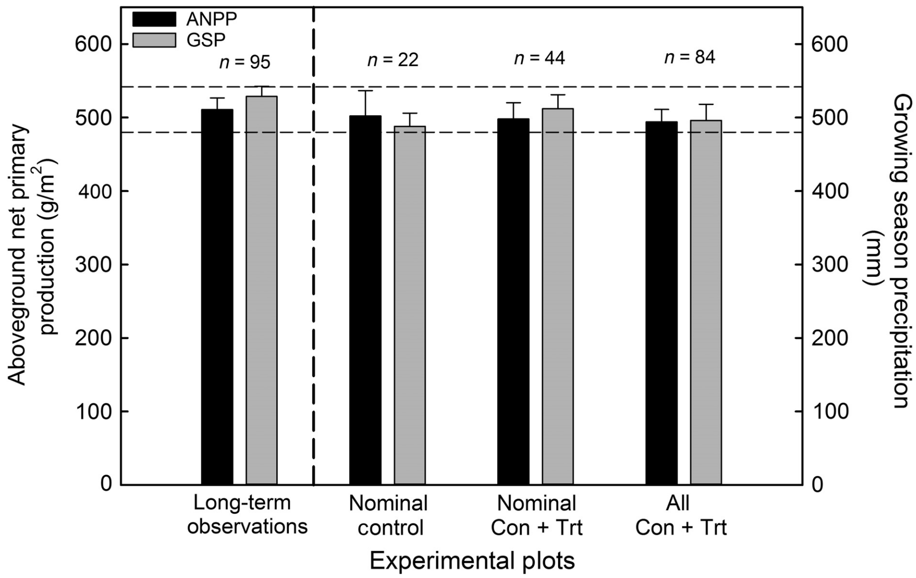

Fig. 2. Mean aboveground net primary productivity (ANPP) and growing season precipitation (GSP; 1 April–31 August) based on all observational data and from three combinations of experimental data (control plots, and control plus treatment plots [Con + Trt] in the nominal range of GSP, and all experimental data regardless of GSP; see Methods for details). The horizontal dashed lines depict the 95% CI for observational ANPP, and the sample size for each estimate is shown above the bars. All estimates of ANPP and GSP were statistically indistinguishable ($P > 0.05$).

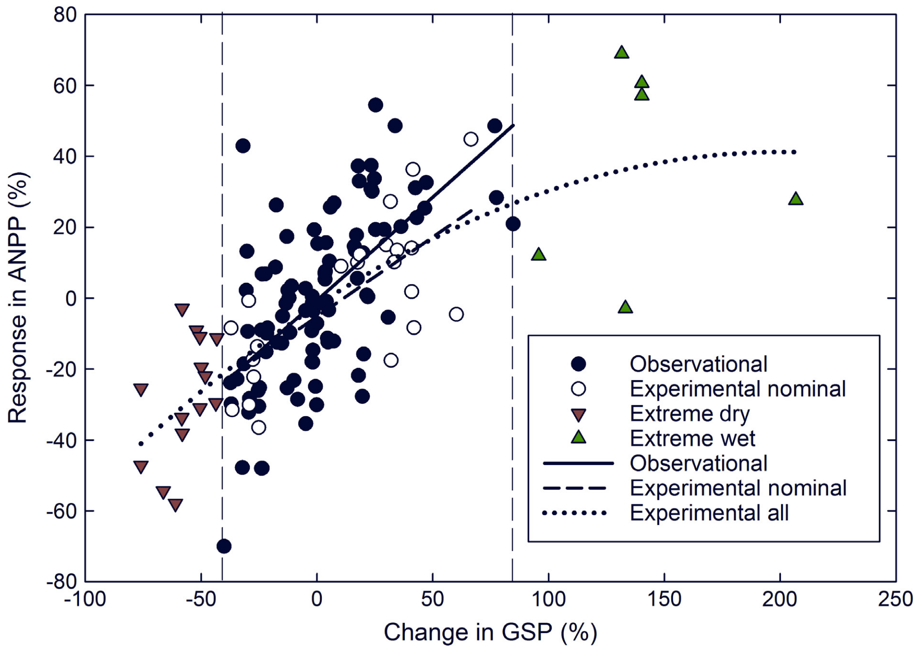

Fig. 3. Relationship between the change in aboveground net primary production (ANPP) in response to growing season precipitation (GSP; 1 April–31 August) for observational data and precipitation experiments conducted at Konza Prairie (see Table 1). Three relationships are shown: the ANPP–GSP relationship for long-term observational data, the relationship for experimental data when GSP was within the nominal range of GSP (denoted by vertical dashed lines), and the relationship for all experimental data including treatments that exceeded nominal levels of GSP (dry or wet). Slopes for the two nominal regressions did not differ significantly ($F = 1.679$, df = 1,113, $P = 0.333$), and the nonlinear relationship was selected based on AIC.

et al. 2011, Kreyling and Beier 2013). Our analysis does not support this perspective.

Second, these experiments were able to replicate reality in a relatively mesic grassland (mean annual precipitation = 850 mm) that is only moderately water limited (ANPP is colimited by multiple resources in tallgrass prairie; Knapp et al. 1998, 2001). Arid and semiarid ecosystems, on the other hand, are much more strongly water limited (Huxman et al. 2004). Thus, we would anticipate that single-factor precipitation manipulation experiments would pass a similar reality check in ecosystems that are more arid as well.

The third insight from this analysis is that while the slopes of the relationships between (standardized) ANPP and GSP did not differ between observations and experiments (Fig. 3), the slope for nominal GSP experiments was less steep than for observations. A steeper slope for observations in the nominal GSP range would be expected if wetter years were generally cooler, and drier years were warmer. This is indeed the case at Konza Prairie where growing season temperature and precipitation (acquired from the Konza Prairie LTER data catalog) are negatively correlated ($r = -0.40$, $P = 0.016$; data *available online*).[5] Results from experiments would not include this covariation between temperature and precipitation, thus, experimental wet years would be expected to be on average warmer than naturally wet years, whereas experimental dry years would be cooler than naturally dry years. As both of these would dampen ANPP responses to altered GSP, we assessed temperature as a covariate for ANPP responses to GSP manipulations. No effect of temperature was detected however, perhaps due to the small range of temperatures encountered in these experimental years.

Finally, when extreme GSP manipulations were included in this analysis, a nonlinear (saturating) GSP–ANPP relationship was suggested (Fig. 3). A nonlinear relationship between ANPP and GSP, when precipitation extremes are included, has been predicted previously (Knapp et al. 2017*b*) but the relatively few extreme wet treatments imposed limits on our ability to support this prediction. This highlights the need for more experiments that simulate extreme wet and dry years, in combination with other predicted global changes, to provide insight into how ecosystems will respond to future changes in precipitation regimes (Kreyling and Beier 2013).

In summary, although results from this "reality check" were positive for this ecosystem, assessments of global change experiments in a wider range of ecosystems, and more experiments with extreme precipitation treatments, are clearly needed. Given the prominent role of experiments in ecology, their deployment along environmental gradients (e.g., Wu et al. 2012), and the availability of long-term data, we should be evaluating much more broadly what can and cannot be learned from experimental approaches.

Acknowledgments

Support was provided by US National Science Foundation funded projects: The Konza Prairie LTER program, The Extreme Drought in Grasslands (EDGE) and Integrating NEON ANPP Data with Existing Long-term Data projects (Macrosystems Biology), Drought-Net (a Research Coordination Network), and The RaMPs project (Ecosystem Science, LTREB).

Literature Cited

Barner, A. K., K. E. Coblentz, S. D. Hacker, and B. A. Menge. 2018. Fundamental contradictions among observational and experimental estimates of non-trophic species interactions. Ecology 99:557–566.

Beier, C., et al. 2012. Precipitation manipulation experiments—challenges and recommendations for the future. Ecology Letters 15:899–911.

Blume-Werry, G., J. Kreyling, H. Laudon, and A. Milbau. 2016. Short-term climate change manipulation effects do not scale up to long-term legacies: effects of an absent snow cover on boreal forest plants. Journal of Ecology 104:1638–1648.

Briggs, J. M., and A. K. Knapp. 1995. Interannual variability in primary production in tallgrass prairie: climate, soil moisture, topographic position and fire as determinants of aboveground biomass. American Journal of Botany 82:1024–1030.

Briggs, J. M., and A. K. Knapp. 2001. Determinants of C_3 forb growth and production in a C_4 dominated grassland. Plant Ecology 152:93–100.

Cherwin, K., and A. K. Knapp. 2012. Unexpected patterns of sensitivity to drought in three semi-arid grasslands. Oecologia 169:845–852.

Copeland, S. M., S. P. Harrison, A. M. Latimer, E. I. Damschen, A. M. Eskelinen, B. Fernandez-Going, M. J. Spasojevic, B. L. Anacker, and J. H. Thorne. 2016. Ecological effects of extreme drought on Californian herbaceous plant communities. Ecological Monographs 86:295–311.

De Boeck, H. J., et al. 2015. Global change experiments: challenges and opportunities. BioScience 65:922–931.

Denton, E. M., J. D. Dietrich, M. D. Smith, and A. K. Knapp. 2017. Aboveground productivity is more sensitive to timing of drought than belowground productivity in a mesic grassland. Plant Ecology 218:317–328.

Elmendorf, S. C., et al. 2015. Experiment, monitoring, and gradient methods used to infer climate change effects on plant communities yield consistent patterns. Proceedings of the National Academy of Sciences USA 112:448–452.

Estiarte, M., et al. 2016. Few multi-year precipitation-reduction experiments find a shift in the productivity-precipitation relationship. Global Change Biology 22:2570–2581.

Fay, P. A., J. D. Carlisle, A. K. Knapp, J. M. Blair, and S. L. Collins. 2003. Productivity responses to altered rainfall patterns in a C4-dominated grassland. Oecologia 137:245–251.

Hoover, D. L., A. K. Knapp, and M. D. Smith. 2014. Resistance and resilience of a grassland ecosystem to climate extremes. Ecology 95:2646–2656.

Hoover, D. L., K. R. Wilcox, and K. E. Young. 2018. Experimental droughts with rainout shelters: a methodological review. Ecosphere 9. https://doi.org/10.1002/ecs2.2088

Huxman, T. E., et al. 2004. Convergence across biomes to a common rain-use efficiency. Nature 429:651–654.

Ives, A. R., J. Foufopoulos, E. D. Klopfer, J. L. Klug, and T. M. Palmer. 1996. Bottle or big-scale studies: how do we do ecology? Ecology 77:681–685.

Knapp, A. K. 1984. Water relations and growth of three grasses during wet and drought years in a tallgrass prairie. Oecologia 65:35–43.

Knapp, A. K., J. M. Briggs, D. C. Hartnett, and S. L. Collins (eds.). 1998. Grassland dynamics: long-term ecological research in tallgrass prairie. Oxford University Press, New York, New York, USA.

Knapp, A. K., J. M. Briggs, and J. K. Koelliker. 2001. Frequency and extent of water limitation to primary production in a mesic temperate grassland. Ecosystems 4:19–28.

Knapp, A. K., J. M. Briggs, and M. D. Smith. 2012*a*. Community stability does not preclude ecosystem sensitivity to chronic resource alteration. Functional Ecology 26:1231–1233.

Knapp, A. K., et al. 2012*b*. Past, present, and future roles of long-term experiments in the LTER network. BioScience 62:377–389.

Knapp, A. K., et al. 2017*a*. Pushing precipitation to the extremes in distributed experiments: recommendations for simulating wet and dry years. Global Change Biology 23:1774–1782.

[5] http://lter.konza.ksu.edu/data

Knapp, A. K., P. Ciais, and M. D. Smith. 2017*b*. Reconciling inconsistencies in precipitation-productivity relationships: implications for climate change. New Phytologist 214:41–47.

Kreyling, J., and C. Beier. 2013. Complexity in climate change manipulation experiments. BioScience 63:763–767.

Leuzinger, S., Y. Luo, C. Beier, W. Dieleman, S. Vicca, and C. Körner. 2011. Do global change experiments overestimate impacts on terrestrial ecosystems? Trends in Ecology and Evolution 26:236–241.

Liu, D., J. Peñuelas, R. Ogaya, M. Estiarte, K. Tielbörger, F. Slowik, X. Yang, and M. C. Bilton. 2018. Species selection under long-term experimental warming and drought explained by climatic distributions. New Phytologist 217:1494–1506.

Luo, Y., et al. 2011*a*. Coordinated approaches to quantify long-term ecosystem dynamics in response to global change. Global Change Biology 17:843–854.

Luo, Y., K. Ogle, C. Tucker, S. Fei, C. Gao, S. LaDeau, J. S. Clark, and D. S. Schimel. 2011*b*. Ecological forecasting and data assimilation in a data-rich era. Ecological Applications 2:1429–1442.

Mayor, J. R., et al. 2017. Elevation alters ecosystem properties across temperate treelines globally. Nature 542:91–95.

Metz, J., and K. Tielbörger. 2016. Spatial and temporal aridity gradients provide poor proxies for plant–plant interactions under climate change: a large-scale experiment. Functional Ecology 30:20–29.

Park, T. 1961. An ecologist's view. Bulletin of the Ecological Society of America 42:4–10.

R Core Team. 2018. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/

Roush, W. 1995. When rigor meets reality. Science 269:313–315.

Saccone, P., and R. Virtanen. 2016. Extrapolating multi-decadal plant community changes based on medium-term experiments can be risky: evidence from high-latitude tundra. Oikos 125:76–85.

Sala, O. E., L. A. Gherardi, L. Reichmann, E. Jobbagy, and D. Peters. 2012. Legacies of precipitation fluctuations on primary production: theory and data synthesis. Philosophical Transactions of the Royal Society of London B 367:3135–3144.

Sandel, B., and A. B. Smith. 2009. Scale as a lurking factor: incorporating scale-dependence in experimental ecology. Oikos 118:1284–1291.

Sandel, B., L. J. Goldstein, N. J. Kraft, J. G. Okie, M. I. Shuldman, D. D. Ackerly, E. E. Cleland, and K. N. Suding. 2010. Contrasting trait responses in plant communities to experimental and geographic variation in precipitation. New Phytologist 188:565–575.

Silvertown, J., et al. 2010. Environmental myopia: a diagnosis and a remedy. Trends in Ecology and Evolution 25:556–561.

Smith, M. D. 2011. An ecological perspective on extreme climatic events: a synthetic definition and framework to guide future research. Journal of Ecology 99:656–663.

Smith, M. D., A. K. Knapp, and S. L. Collins. 2009. A framework for assessing ecosystem dynamics in response to chronic resource alterations induced by global change. Ecology 90:3279–3289.

Strauss, S. Y., J. A. Lau, T. W. Schoener, and P. Tiffin. 2008. Evolution in ecological field experiments: implications for effect size. Ecology Letters 11:199–207.

Thompson, R. M., J. Beardall, J. Beringer, M. Grace, and P. Sardina. 2013. Means and extremes: building variability into community-level climate change experiments. Ecology Letters 16:799–806.

Tilman, D. 1989. Ecological experimentation: strengths and conceptual problems. Pages 136–157 *in* G. E. Likens, editor. Long-term studies in ecology. Springer, New York, New York, USA.

VanderWeide, B. L., D. C. Hartnett, and D. L. Carter. 2014. Belowground bud banks of tallgrass prairie are insensitive to multi-year, growing-season drought. Ecosphere 5:1–17.

Wardle, D. A. 2016. Do experiments exploring plant diversity–ecosystem functioning relationships inform how biodiversity loss impacts natural ecosystems? Journal of Vegetation Science 27:646–653.

Wilcox, K. R., J. C. von Fischer, J. M. Muscha, M. K. Petersen, and A. K. Knapp. 2015. Contrasting above- and belowground sensitivity of three Great Plains grasslands to altered rainfall regimes. Global Change Biology 21:335–344.

Wilcox, K. R., J. M. Blair, M. D. Smith, and A. K. Knapp. 2016. Does ecosystem sensitivity to precipitation at the site-level conform to regional-scale predictions? Ecology 97:561–568.

Williams, J. W., and S. T. Jackson. 2007. Novel climates, no-analog communities, and ecological surprises. Frontiers in Ecology and the Environment 5:475–482.

Wolkovich, E. M., et al. 2012. Warming experiments underpredict plant phenological responses to climate change. Nature 485:494–497.

Wu, Z., P. Dijkstra, G. W. Koch, J. Penuelas, and B. A. Hungate. 2011*a*. Responses of terrestrial ecosystems to temperature and precipitation change: a meta-analysis of experimental manipulation. Global Change Biology 17:927–942.

Wu, Z., G. W. Koch, P. Dijkstra, M. A. Bowker, and B. A. Hungate. 2011*b*. Responses of ecosystem carbon cycling to climate change treatments along an elevation gradient. Ecosystems 14:1066–1080.

Wu, Z., P. Dijkstra, G. W. Koch, and B. A. Hungate. 2012. Biogeochemical and ecological feedbacks in grassland responses to warming. Nature Climate Change 2:458–461.

Yuan, Z. Y., F. Jiao, X. R. Shi, J. Sardans, F. T. Maestre, M. Delgado-Baquerizo, P. B. Reich, and J. Peñuelas. 2017. Experimental and observational studies find contrasting responses of soil nutrients to climate change. eLife 6:e23255.