ECOLOGY LETTERS

Ecology Letters, (2021) **24**: 1103–1111 doi: 10.1111/ele.13710

REVIEW AND SYNTHESIS

How long do population level field experiments need to be? Utilising data from the 40-year-old LTER network

Sarah Cusser,[1,2]*
Jackson Helms IV,[1,3]
Christie A. Bahlai[1,4] and
Nick M. Haddad[1]

Abstract

We utilise the wealth of data accessible through the 40-year-old Long-Term Ecological Research (LTER) network to ask if aspects of the study environment or taxa alter the duration of research necessary to detect consistent results. To do this, we use a moving-window algorithm. We limit our analysis to long-term (> 10 year) press experiments recording organismal abundance. We find that studies conducted in dynamic abiotic environments need longer periods of study to reach consistent results, as compared to those conducted in more moderated environments. Studies of plants were more often characterised by spurious results than those on animals. Nearly half of the studies we investigated required 10 years or longer to become consistent, where all significant trends agreed in direction, and four studies (of 100) required longer than 20 years. Here, we champion the importance of long-term data and bolster the value of multi-decadal experiments in understanding, explaining and predicting long-term trends.

Keywords

Data mining, isothermality, long-term, moving window, population dynamics, time series, trajectory.

Ecology Letters (2021) **24**: 1103–1111

INTRODUCTION

Long-term experiments are essential in the study of ecology: critical in isolating and understanding the ecological consequences of global land use and climate change (Likens *et al.* 1996; Del-Val & Crawley 2005; Haddad *et al.* 2015; Gonzalez *et al.* 2016; Hughes *et al.* 2017; Van Klink *et al.* 2020). Long-term data, collected over at least 10 years, are especially important in evaluating ecosystem properties and processes that require evaluation over a full range of regional climatic extremes or are slow to develop over time (Tilman *et al.* 1994; Rasmussen *et al.* 1998; Knapp *et al.* 2012). However, for a variety of reasons, short-term experiments lasting 1–5 years are the benchmark in ecology. Short-term experiments, which are more consistent with typical grant cycles and graduate programs, are essential in identifying ecosystem-related changes in a timely and cost-effective manner. Despite their importance, conceptual problems have occurred because ecologists use short-term experiments to address long-term questions (Tilman 1989). Temporally restricted research may merely capture a snapshot of ecosystem properties as they gradually respond to manipulation (Hanski & Ovaskainen 2002; Helm *et al.* 2006; Knapp *et al.* 2012; Jarvis & Williams 2016; Voelkl & Würbel 2016). As such, research conducted at constrained time scales has the potential to be misleading, either capturing short-term trends or failing to detect trends at all (Bahlai *et al.* 2020; Cusser *et al.* 2020). While the argument for long-term data collection is not new (Callahan 1984; Tilman 1989; Tilman 1997; Rees *et al.* 2001; Estes *et al.* 2018, among others), rarely have multiple datasets been amassed at decadal timescales appropriate to instill confidence in proposed long-term trends. One place where this is possible, and is the focus of our study, is in the 40-year-old Long-term Ecological Research (LTER) network.

Developed in direct response to arguments for the importance of long-term experimental studies (Callahan 1984), the LTER network not only provides a 'sandbox' in which to examine long-term responses to experimental manipulation, it also allows us to contextualise hypothetical shorter term studies by parsing apart ephemeral, lagged or spurious responses from those that are genuine changes in system behaviour. While previous research has suggested the importance of long-term data, never before has the spatial or temporal extent and quality of long-term data been available to examine its importance, until now.

Ecological systems are inherently dynamic, and variation is driven by a variety of stochastic and deterministic processes (Folke 2006; Suding & Gross 2006; Hastings 2010; Beckage *et al.* 2011; McCann *et al.* 2020). Confident experimental outcomes in highly dynamic abiotic environments are likely to require evaluation over an appropriately variable climate. Ideally, a study should capture the full range of abiotic extremes representative of that system (Ives & Carpenter 2007). The longer that an experiment proceeds, the more likely its results will reflect both the average and extreme climatic conditions that characterise a particular geographic region (Tilman

---

[1] *W.K. Kellogg Biological Station, Department of Integrative Biology, Michigan State University, Hickory Corners,MI, USA*

[2] *Rubenstein School of Environment and Natural Resources, University of Vermont, Burlington, VT, USA*

[3] *USDA-ARS, Center for Medical, Agricultural, and Veterinary Entomology, Gainesville, FL, USA*

[4] *Department of Biological Sciences, Kent State University, Kent, OH, USA*

**Correspondence: E-mail: sarah.cusser@gmail.com*

1989). Beyond environmental variables, specific taxa under investigation may respond slowly to experimental manipulation, only reaching a delayed response after some temporal threshold is met (Krauss *et al.* 2010). These temporal thresholds are likely to be closely linked to taxa-specific life-history traits, including generation time, dispersal and colonisation ability, and dormancy periods, among others. For example if long-lived plants can survive initial experimental disruption, changes in plant population abundance may take many generations to become apparent, even if the immediate results are measurable in reduced individual fitness (Tilman *et al.* 1994; Cousins *et al.* 2007; Ellis & Coppins 2007; Gustavsson *et al.* 2007; Jackson *et al.* 2009; Haddad *et al.* 2015). Also, a plant's seed bank may further prolong the lag in response, replacing individuals lost in the adult population following disturbance (Plie *et al.* 2017). Some animals, due to their relatively short generation time, high mobility, and potential to track resources in novel environments, may respond more rapidly to manipulation (Kuussaari *et al.* 2009; Krauss *et al.* 2010), and may consequently not require long experiments to confidently determine consistent results from manipulation.

We focus our quantitative synthesis on a single response type in experimental studies: population level organismal abundance. While patterns of abundance are themselves a fundamental topic in ecology, they also underlie some of the most basic questions in the field and have been used to develop hypotheses concerning species response to climate change, identify probable locations of pest outbreaks, and choose the location of natural reserves (Elton & Nicholson 1942; Altieri *et al.* 1984; Pounds *et al.* 1999; Sagarin *et al.* 2006). Given that measures of organismal abundance are relatively quick to execute, consistently apparent between observers and years, and an intuitive measure of population condition in some systems, abundance is a regularly collected and relatively comparable metric between studies.

Here we make two hypotheses concerning the importance of long-term studies (> 10 years) in documenting organismal abundance: (H1) If studies take place in highly variable environments, with increased system-specific abiotic variation, then studies of those systems will require longer periods to detect consistent results than those studies in environments with more consistent abiotic variables. (H2) If taxa have long generation times or low dispersal and colonisation abilities, then studies of those taxa will result in a higher proportion of spurious short-term trends than taxa with shorter generation times, high mobility and potential to track resources in novel environments. To test these hypotheses, we use a moving window algorithm and utilise the wealth of data across studies of organismal abundance mined from the 40-year-old Long-Term Ecological Research (LTER) network. We posit that long-term experiments may be needed to understand, explain and predict long-term trends.

METHODS

Data mining

We searched the 40-year-old Long-term Ecological Research database network portal (https://portal.edirepository.org/nis/home.jsp) to identify and repurpose relevant long-term experimental datasets. We systematically explored each of the 6957 unique datasets, from 30 locations that were available as of December 2018. Only datasets that met five requirements were included in our analysis: (1) research spanned ten years or longer; (2) research recorded data in ten distinct years and data could be expressed as a summary metric for each year of research (i.e. average organismal abundance by treatment); (3) research documented a press experiment in which the treatment was repeatedly applied throughout the experimental period (Bender *et al.* 1984) and treatments could be divided into a 'control' and 'treatment' category; (4) treatment response was recorded as a measure of organismal abundance and (5) research showed evidence of change in treatment effect over time (i.e. included at least one statistically significant linear relationship between the effect size and time, discussed below). Some study sites recorded multiple datasets documenting organismal abundance, and some datasets quantified multiple taxa responses to the same experimental manipulation. Time series were divided into the finest taxonomic resolution available for analysis (i.e. order, genera, species, or morphospecies). Whenever possible, each taxon within each dataset at each site was analysed separately. A few of the datasets were resampled multiple years after a continuous sampling effort, resulting in a single, temporally disconnected, sampling point. To ensure continuity with other, consistently sampled studies, these single 'late-breaking' datapoints were removed from analysis.

Our search identified 100 datasets from 28 distinct studies and 12 LTER sites that met our five requirements (Fig. 1, ESM Table 1). Another 22 datasets met our first four requirements, but lacked any significant linear relationship between the effect size and time. In these datasets, the effect size did not change significantly over the course of study regardless of the study duration. As such, they were not likely to be of interest for long-term analysis. Before the removal of these 22 datasets, we analysed all 122 time-series that met the first four requirements. The results followed similar patterns with and without the 22 datasets, thus we felt confident in our decision to remove them from the analysis.

The experimental scope of the 100 datasets used for analysis ranged from studies on the exclusion of herbivores [Sevilleta (Lightfoot, 2010, 2016a, 2016b, 2016c) and Short Grass Steppe (Milchunas 2014)], to manipulating moisture [Konza Prairie (Smith *et al.* 2020; Blair 2020; Joern 2020) and Sevilleta (Pockman 2013; Collins 2016b)], nutrients [Arctic (Shaver 2016), Cedar Creek (Tilman 2018c), Hubbard Creek (Hamburg 2016), and Plum Island (Deegan 2012; Deegan & Warren 2012)], pH [North Temperate Lakes (Kratz 2019)] and temperature [McMurdo Dry Valleys (Wall & Virginia 2016)], as well as deliberately altering species diversity [Cedar Creek (Tilman 2018a, 2018b)], or removing plants by fire, grazing [Konza Prairie (Boyle 2020)], trimming [Luquillo (Sharpe 2017), Santa Barbara (Reed & Harrer 2017; Santa Barbara LTER & Reed 2020a, 2020b)], mechanical tillage [Kellogg (Landis 2018)] or some combination thereof [Sevilleta (Collins *et al.* 2016; Collins 2016a)]. For a full list of LTER sites and experiments involved in our analyses see ESM Table 1.

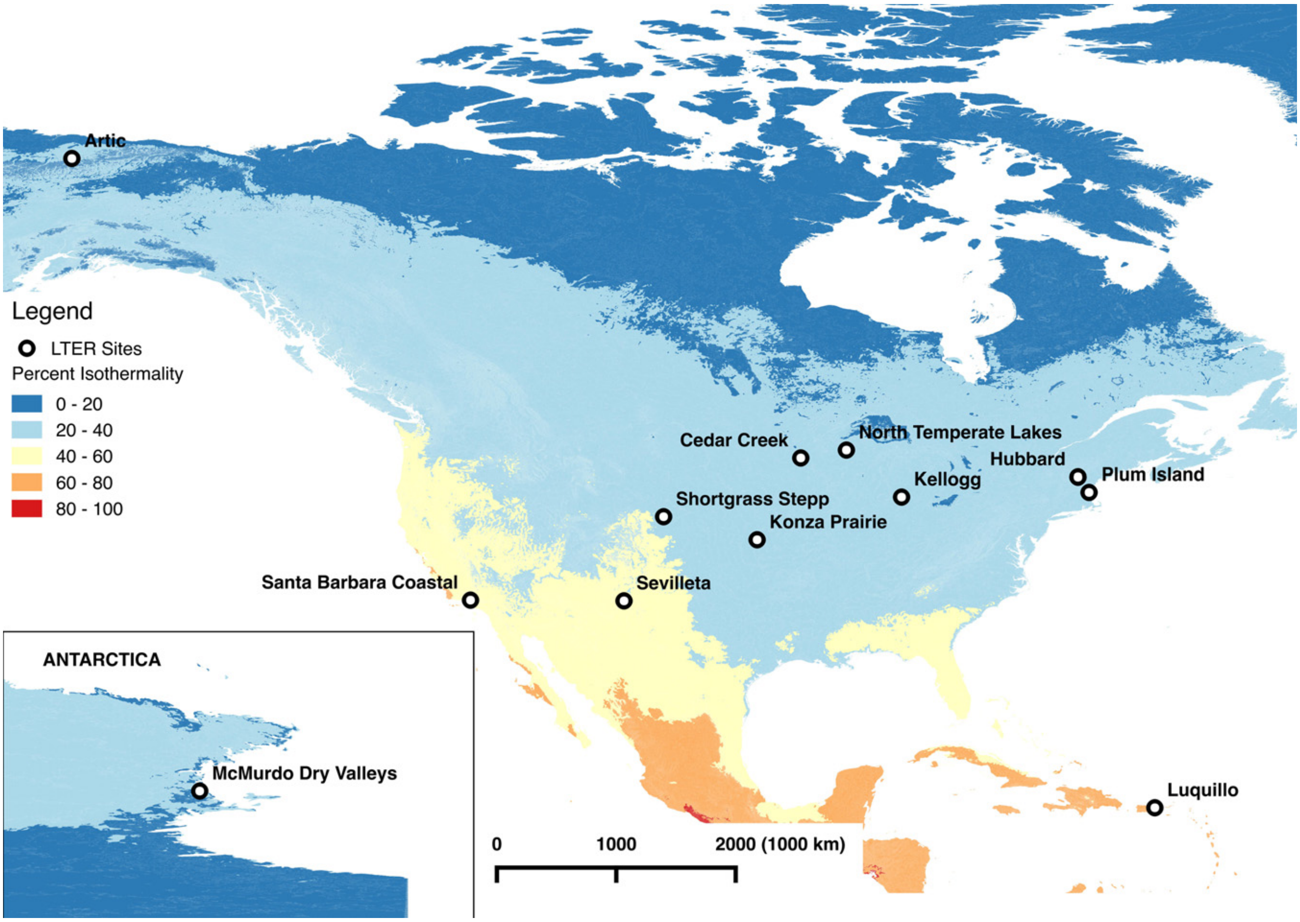

Figure 1 Map showing 12 LTER sites with data in our study across North America and in Antarctica (inset). Colours represent climate variability as determined by isothermality. (BIOCLIM Variable 3). Lower isothermality (cooler colours) indicate higher annual climate variability. Higher isothermality (warmer colours) indicate lower annual climate variability.

Effect size and moving window algorithm

To determine the study duration needed to identify long-term trends, we first calculated the effect size of the press treatment for each year of each of the 100 datasets as Hedges' *g*. As such, effect size was calculated as: $[x_t - x_c]/SD_p$, where x_t is the average treatment population size in that year, x_c is the average control population size in that year and SD_p is the pooled standard deviation of that year (Rosenthal *et al.* 1994).

Next, to understand the study duration needed to identify changes in effect size over time for each dataset, we applied a moving window algorithm developed in R (Bahlai *et al.* 2020). We fit linear regression models to defined subsets of each dataset. Datasets were divided into every possible subset of three consecutive year periods or longer. For example a dataset of 10 years would include analysis of a total of 36 window subsets: 8 distinct three-year windows, 7 four-year windows, 6 five-year windows and so on, with the final window encompassing the full dataset, or single 10-year window. From each window subset, we separately analysed a linear regression and compiled summary statistics of interest (i.e. slope of the relationship between Hedges' g and time, standard error of this relationship and p-value). See Bahlai *et al.* 2020 for an in-depth description of the process and github (https://github.com/cbahlai/broken_window) for the appropriate R code.

We defined the significance and direction of the longest time series (i.e. the slope of the linear regression of the whole dataset) as a proxy for the 'true' trajectory of the data, as it represents the best information available. Significance was determined at the alpha = 0.05 level. Comparisons of the data subsets are based on comparison to these 'true' trends and were consequently sorted into two groups: those with long-term trends (i.e. those with a significant relationship between change in effect size and time for the entire dataset, Fig. 2a) and those without (i.e. those with no significant relationship between change in effect size and time for the entire dataset, Fig. 2b). Conceptually, we were interested in the trajectory of the relationship between Hedges' g and time, and how linear regression model outputs vary with sample period duration: our focus was determining the probability of patterns being observed. Thus, our methods are not to form specific ecological conclusions concerning experiments and we make no claims about the ecological implications of the trends we found. Rather, we

Table 1 (a) Parameter estimates of generalised linear mixed models (GLMMs) to determine the relationship between isothermality, the categorical variables (plant or animal), as well as any interactions between the two with critical temporal threshold. We use 'isothermality' and 'plant/animal' as fixed effects and 'LTER dataset' nested within 'LTER Site' as a random intercept. The continuous climatic variable was scaled to account for differences in magnitude. We tested for overdispersion, of which we found no evidence. (b) Linear regression model to test for significant differences in percent spurious trends between studies focused on plants and those on animals

(a) Critical temporal threshold

Fixed effects	Estimate	SE	*z* value	*P* value
Intercept (Animal)	2.06	0.12	17.75	<0.001
Isothermality	0.06	0.11	0.52	0.6
Plant/Animal	0.48	0.15	3.24	0.001
Isothermality × Plant	−0.44	0.144	-3.21	0.001

(b) Percent spurious results

Fixed effects	d.f.	Sum Sq	Mean Sq	*F* value	*P* value
Plant/Animal	1	0.0626	0.06265	4.875	0.029
Residuals	98	1.2594	0.01285		

encourage that any dataset-specific results be interpreted with a degree of caution.

No adjustments were made for multiple statistical comparisons in our analysis as each linear regression was considered in isolation, as a hypothetical observation period which an experimenter would use to reach conclusions regarding system behaviour, from non-independent but still separate experimental durations.

Extracting temporal thresholds and percent spurious trends

To quantify the necessary study duration for each dataset, we plotted window length (i.e. number of years) in each subset of the moving window algorithm against the value of the slope for that window's linear regression. The result was 100 distinct 'pyramid plots' (Bahlai *et al.* 2020; Cusser *et al.* 2020). From each plot, we extracted a value for 'critical temporal threshold'. For studies with long-term trends (Fig 2a), a critical temporal threshold is determined as the minimum number of years until all trends agreed with the long-term trend (i.e. all trends of that window length are significant and in the same direction as the long-term trend). For each dataset lacking a long-term trend (Fig 2b), the critical temporal threshold is the minimum number of years to avoid all spurious results (i.e. all trends of that duration are not significant thus avoiding misleading significant results). As such, we recorded a value for critical temporal threshold for both datasets with and without long-term trends.

We also calculated the percent of spurious results found in each of the 100 datasets. For datasets with 'true' long-term trends, spurious trends were those periods with significant slopes in the direction opposite to the long-term (Fig 2a). For datasets which lacked long-term trends, any significant results were considered spurious (Fig 2b). Percent spurious trends were determined for each study as the number of spurious trends/total number of trends.

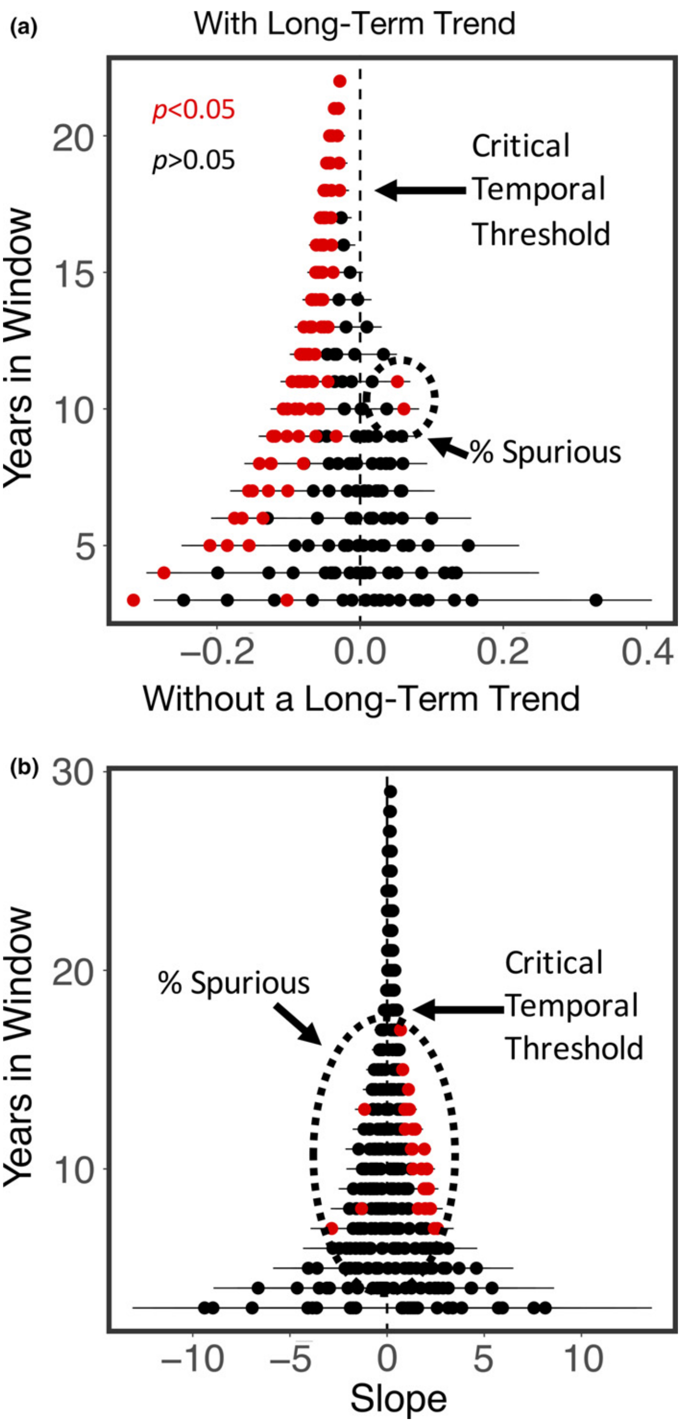

Figure 2 Example moving window plots showing critical temporal thresholds and spurious results from a dataset with a long-term trend (a) and dataset without a long-term trend (b). Each plot represents a single experimental study tracking organismal abundance. Red dots show significant trends at the alpha = 0.05 level. Black dots represent non-significant trends. Positive regression slopes indicate that organismal abundance increased in the control relative to treatment, whereas negative slopes indicate the opposite. Panel (a) shows data from the Konza Prairie LTER (knz.72.8) *Andropogon gerardii* response to moisture manipulation. The percent spurious trends are the number of red dots with positive slope by the total number of dots. Panel (b) also shows data from the Konza Prairie LTER (knz.26.10) *Dickcissel* response to plant removal by fire. The percent spurious trends are the number of red dots divided by the total number of total dots.

Continuous and categorical explanatory variables

To address our first hypothesis, concerning the influence of abiotic variation on study duration, we extracted the WorldClim v2 Bioclim variable of Isothermality for each of the 12 LTER locations (Busby 1991). Initially, we had explored multiple Bioclim variables, including precipitation seasonality and annual precipitation, but variables were determined to be either collinear or uninformative and were consequently removed from further analysis. As a result, we use the BIOCLIM variable of isothermality as a single measure of intra-annual temperature fluctuation at each of our sites. Isothermality quantifies how large the day-to-night temperatures oscillate relative to the summer- to-winter (annual) oscillations (Isothermality = (Mean Diurnal Temperature range/Temperature annual range) × 100). As such, sites with low isothermality are located in the most variable abiotic environments. Across datasets, isothermality averaged 37.65% (SE: 1.22%) and ranged from 21.44% (McMurdo Dry Valleys) to 71.52% (Luquillo).

To address our second hypothesis, concerning the extent of spurious results surrounding slow to change taxa, we coarsely divided datasets based on taxa specific life-history traits, including generation time, dispersal and colonisation ability and dormancy periods. As a proxy for these difficult to measure life-history traits, we divided datasets into two categories: those focused on plants and those focused on animals. Of the 100 time-series, 56 focused on animal abundance and 44 on plant abundance.

Analysis

To address our first prediction, that studies taking place in highly variable abiotic environments will require longer study periods to reach consistent results than more abiotically stable systems, we use generalised linear mixed models (GLMMs) to determine the relationship between isothermality and our categorical explanatory variable (i.e. plant or animal), as well as any interactions between the two with our response (i.e. critical temporal threshold) using the 'glmer' function in the package 'lme4' (Bates *et al.* 2014). We used 'isothermality' and 'plant/animal' as fixed effects and 'LTER dataset' nested within 'LTER Site' as a random intercept. We scaled continuous climatic variables to account for differences in magnitude. We tested for overdispersion, of which we found no evidence.

To address our second prediction, that experiments investigating slow to change (i.e. plant) taxa will be more often characterised by high proportions of misleading short-term trends than experimental studies focused on quick to respond (i.e. animal) taxa, we used linear regression to test for significant differences in percent spurious results between studies that focused on plants and those on animals.

RESULTS

Of the 100 datasets, we found 24 studies with a significant trend for the full dataset (change in effect size over the full study period), and 76 studies without a significant trend for the full dataset. It took 9.66 years on average (SE: 0.52, range 3–32 years) to reach a critical temporal threshold and achieve consistent significant results. On average, 11.7% (SE: 1.1%, range 0.7–47%) of significant trends derived from subsets of sampling years were spurious, not agreeing with the long-term pattern of the data.

We found support for our predictions, finding that both isothermality and the plant/animal distinction contributed to the length of the critical temporal threshold necessary for consistent results. We found that studies taking place in highly variable abiotic environments required the longest periods of study to reach consistent results. As such, those sites located in the most dynamic abiotic environments (those with low isothermality) required the longest periods of evaluation. On average, we see that for every 1% increase in abiotic variation (1% decrease in isothermality), we saw approximately a 0.1-year (1.2 months) extension of the critical temporal threshold across taxa (Fig. 3a). We also found an interaction between isothermality and our taxa variable (plant or animal) in explaining the length of time needed to reach consistent results (Estimate: −0.44, SE = 0.14, $z = -3.20$, $P = 0.001$, Fig. 3a, Table 1). Plant studies required longer periods of time to become consistent in highly dynamic environments. For example studies undertaken at the Cedar Creek and Arctic LTERs, which are plant studies characterised by the strongest seasonal extremes in our study, also had the longest critical temporal thresholds (32 and 16 years respectively).

We also found that studies focused on plants were characterised by significantly more spurious results than those that focused on animals ($T = 2.2$, $P = 0.03$, Fig 3b). On average, 15% of significant plant trends were spurious, compared to only 9% of animal trends. We also show that every study investigated contained at least one spurious trend, and most studies (63%) had more than the expected number of false positives, or type I error, expected at the traditional 0.05 alpha threshold.

DISCUSSION

We found support for both of our predictions: experimental studies in dynamic abiotic environments generally needed longer periods of study than those in more stable environments. We also found that experiments investigating potentially slow to change taxa, such as plants, were more often characterised by misleading short-term trends than those studies focused on animals. Most importantly, we underscore the importance of using long-term data to address long-term questions. We see that nearly half (46/100) of the studies we investigated require 10 years or longer to reach stable and consistent results, and four studies required longer than 20 years.

While our unprecedented access to long-term data broadens both the scope and magnitude of our findings, the themes we posit are by no means revolutionary. Ecologists have long recognised that while short-term experiments offer important insights into transient dynamics, transient dynamics may not reflect long-term trends (Callahan 1984; Tilman 1989). For example the Park Grass Experiments of Rothamsted, England underscore the importance of long-term data. As one of the longest of all ecological experiments to date, Park Grass

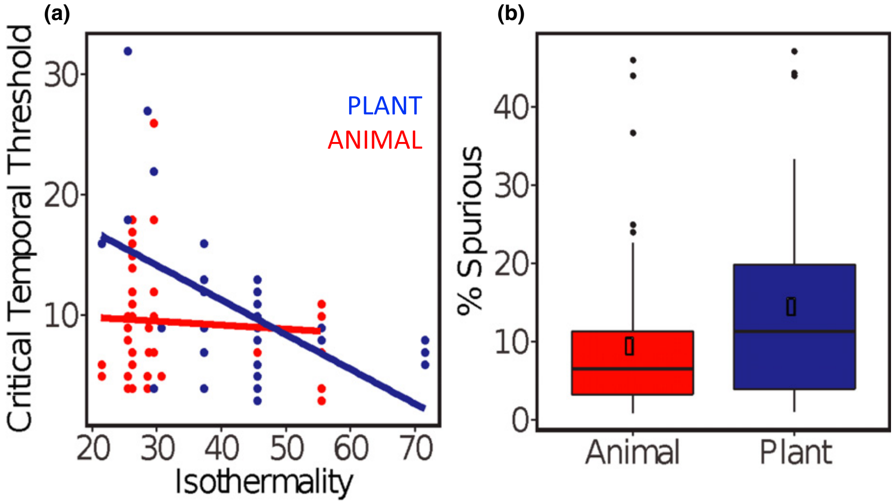

Figure 3 (a) Relationship between isothermality (%) and critical temporal threshold (years), which is more apparent in plants (blue) than in animals (red). (b) Boxplot comparing percent spurious results between studies of animals and plants. The central bar gives group median, boxes give the 1st and 3rd quartiles, closed circles show outliers and open circles show group mean.

investigates how research conclusions may vary depending on study length (Lawes & Gilbert 1880; Brenchley & Warington 1958; Tilman 1982). The Park Grass plots, which were subjected to a repeated pattern of nutrient addition for over 130 years, show that nutrient dynamics are extremely slow to develop, especially in direct comparison to studies on resource competition. Authors conclude that 5-year studies are likely too short to understand long-term dynamics. While similar in motivation and outcome to our work, the Park Grass Experiments investigate a single experimental manipulation within one region. Our analysis incorporates a large spatial scale, transcending biogeographic regions, and includes multiple treatment types and a range of response taxa to show that patterns found in the Park Grass Experiment are more universal than initially claimed.

Beyond the population level press experiments that we study here, the LTER boasts a range of observational and experimental research. Pulsed experimental treatments, which mimic one-time disturbance events or single catastrophic environmental changes, may be a natural extension of our methods. While pulsed manipulations can result in short periods of high productivity following disturbance, they can also generate long-lasting, complex, or lagged impacts (Haddad *et al.* 2002). For example the Hubbard Brook experiment offers an interesting case (Likens *et al.* 1996). During the long monitoring period following a pulse treatment, decades passed before researchers could uncover genuine long-term changes in stream chemistry (Rosi-Marshall *et al.* 2016; Marinos *et al.* 2018). Researchers hypothesise that the slow to emerge effects were related to gradual changes in forest community composition. The Hubbard Brook experiment offers evidence that ecosystem dynamics can be impacted long beyond the occurrence of a major disturbance event. Determining how and when systems respond to pulsed treatments may be an interesting extension of our moving window analysis.

While the focus of this study is on the population level metric of organismal abundance, our technique is readily applicable to higher level community or ecosystem processes. This is particularly important for the LTER network, as the five guiding criteria of the network are at the community (trophic structure) and ecosystem (productivity, organic matter accumulation, nutrients, disturbance) levels. Given their complexity, there is reason to hypothesise that higher-level properties which define community or biogeochemical processes may take even longer to become consistent, and be defined by more spurious trends, than the studies of population level organismal abundance that we investigate here. Establishing temporal trajectory in ecological systems is prone to other pitfalls and biases, including site selection biases, transient dynamics upon the establishment of a study, and sampling effects, and these issues can be magnified in population studies due to density-dependent nonlinear dynamics inherent to these systems (Fournier et al 2019; Wauchope *et al.* 2019; Didham et al 2020).

Owing to high temporal variability in community dynamics and lags in biodiversity response, changes in trophic structure may take decades or longer to be described and explained (Tilman *et al.* 1994; Kuussaari *et al.* 2009; Magurran *et al.* 2010; Record *et al.* 2020). The rate of community response following experimental disturbance is likely to depend on taxa specific traits within that community and consequently likely to play out over longer time periods than any individual species considered in isolation. For example longer-lived species may be lost more slowly than shorter-lived species; lower trophic levels lost more slowly than higher trophic levels (Borrvall & Ebenman 2006; Staddon *et al.* 2010); and generalist species may decline more slowly than specialist species (Watts *et al.* 2020). Furthermore, extinction debt predicts that the most abundant species (i.e. the best competitors) are likely to go extinct first following disturbance (Tilman *et al.* 1994), leaving less competitive (e.g. Tilman 1997) or longer-lived species remaining in the community long after the habitat can no longer support them. Given that the focus of our study is often on the most common species in a community, extending our methods to include less common species may increase the prevalence of extinction debt and lagged responses at the community level. Indeed, models created by Tilman *et al.* (1994) suggest that extinction debt can play out over scores of decades and resulting species interactions (i.e. trophic cascades), affected by order and timing of changes in species abundance and presence, will take decades to develop (Staddon *et al.* 2010).

The order of species loss may affect not only community disassembly, but also the magnitude and trajectory of changes in ecosystem properties (Lawton 1994; Grime 1998; Gonzalez & Chaneton 2002). We expect biodiversity ecosystem function (BEF) relationships to span multiple spatial and temporal scales (Gonzalez *et al.* 2020). With cross-scale feedbacks, the strength and form of ecosystem response to disturbance is likely to vary across the distribution of individuals within and among species (Gonzalez *et al.* 2020). Thus, variation in ecosystem function will respond not just to variation in species presence, community composition and trophic structure; but also, on effects on carbon and nitrogen availability, soil moisture, productivity and other ecosystem characteristics (Staddon *et al.* 2010; Haddad *et al.* 2015; Cusser *et al.* 2020). The likely consequence is a further lag in ecosystem level response.

Given the extent of ongoing global land use and climate change, long-term experiments are more necessary than ever to understand, explain and predict long-term trends. With global climate change increasing abiotic variability worldwide, results from short-term studies may become increasingly unreliable. While long-term experiments are expensive, both in their cost to operate and in the career-long devotion needed to see trends emerge, new efforts should work in parallel, coordinating network-wide experiments and syntheses across ecosystems and climates. Understanding the relationship between transient and long-term dynamics is a significant challenge that ecologists must tackle, and long-term experiments will be essential for relating observation to theory now, as well as in the future.

ACKNOWLEDGEMENTS

The support for this research was also provided by the NSF Long-term Ecological Research Program (DEB 1832042) at the Kellogg Biological Station, the National Science Foundation Directorate for Computer and Information Science and Engineering (OAC 1838807), USDA National Institute on Food and Agriculture, and by Michigan State University AgBioResearch. We are grateful for conversations with Andy Gonzalez and Chris Klausmeier which added great insight into our interpretations. LTER sites with data participating in our analysis were notified and acknowledged the use of their data. Arctic LTER funding is from NSF under Grants # DEB 1637459, 1026843, 0423385, 9810222, 9211775 and 8702328. Cedar Creek LTER funding is from NSF under Grant # DEB 1831944. Hubbard Brook funding is from NSF under Grants # DEB 9211768, 9810221 and 0423259. All the Hubbard Brook data are published in EDI with the CC- by license: https://creativecommons.org/licenses/by/4.0/. Luquillo LTER funding is from NSF under Grant # DEB 0218039 and 0620910 to the Institute for Tropical Forestry, USDA Forest Service, as part of the Luquillo Long-Term Ecological Research Program. McMurdo Dry Valley LTER funding is from NSF under Grant # OPP 1637708. North Temperate Lakes LTER funding is from NSF under Grant # LTREB 9815519 and LTREB 9527669. Plum Island LTER funding is from NSF under Grant # OCE 1637630 as part of the Plum Island Ecosystems Long-Term Ecological Research Project: Dynamics of coastal ecosystems in a region of rapid climate change, sea-level rise and human impacts. Santa Barbara LTER funding is from NSF under Grant # OCE 1232779 and OCE 1831937. Sevilleta funding is from NSF under Grant # DEB 1655499. Short Grass Steppe LTER funding is from NSF under Grant # 1027319.

AUTHORSHIP

SC and JH conceived the idea with encouragement from NMH. CB developed the moving window algorithm. SC and JH gathered datasets. SC ran algorithms, compiled data and executed analyses. All the authors discussed the results and took part in writing the manuscript.

CONFLICT OF INTEREST

The authors have no conflict of interest to declare.

PEER REVIEW

The peer review history for this article is available at https://publons.com/publon/10.1111/ele.13710.

DATA AVAILABILITY STATEMENT

Data are publicly available in the LTER portal (https://portal.edirepository.org/nis/home.jsp).

REFERENCES

Altieri, M.A., Letourneau, D.K. & Risch, S.J. (1984). Vegetation diversity and insect pest outbreaks. *Crit. Rev. Plant. Sci.*, 2, 131–169.

Bahlai, C., White, E., Perrone, J., Cusser, S. & Whitney, K.S. (2020). An algorithm for quantifying and characterizing misleading trajectories in ecological processes. *BioRxiv preprint*, https://doi.org/10.1101/2020.07.07.192211

Bates, D., Mächler, M., Bolker, B. & Walker, S. (2014). Fitting linear mixed-effects models using lme4. *arXiv preprint* arXiv:1406.5823.

Beckage, B., Gross, L.J. & Kauffman, S. (2011). The limits to prediction in ecological systems. *Ecosphere*, 2(11), 1–12.

Bender, E.A., Case, T.J. & Gilpin, M.E. (1984). Perturbation experiments in community ecology: theory and practice. *Ecology*, 65, 1–13.

Blair, J. (2020). WAT01 Konza Prairie long-term irrigation transect study ver 12. Environmental Data Initiative. https://doi.org/10.6073/pasta/f0fd664ab5bc0c68e392cdcec91316a3. Last accessed 11 February 2021.

Borrvall, C. & Ebenman, B. (2006). Early onset of secondary extinctions in ecological communities following the loss of top predators. *Ecol. Lett.*, 9, 435–442.

Boyle, A. (2020). CBP01 Variable distance line-transect sampling of bird population numbers in different habitats on Konza Prairie ver 11. Environmental Data Initiative. https://doi.org/10.6073/pasta/d966d18585729a2a29eead60f4764f14. Last accessed 11 February 2021.

Brenchley, W.E. & Warington, K. (1958). *The Park Grass Plots at Rothamsted.* Harpenden Rothamsted Research, Harpenden.

Busby, J. (1991). BIOCLIM-a bioclimate analysis and prediction system. *Plant Prot. Q.*, 61, 8–9.

Callahan, J.T. (1984). Long-term ecological research. *Bioscience*, 34(6), 363–367.

Collins, S. (2016a). Nitrogen Fertilization Experiment (NFert): Net primary production quadrat data at the Sevilleta National Wildlife Refuge, New Mexico (2004–present). Environmental Data Initiative. https://doi.org/10.6073/pasta/86f3928d4c2a7eb17de93857d1dfec4e. Last accessed 11 February 2020.

Collins, S. (2016b). Monsoon Rainfall Manipulation Experiment (MRME): Net primary production quadrat data at the Sevilleta National Wildlife Refuge, New Mexico (2006–present). Environmental Data Initiative. https://doi.org/10.6073/pasta/987342540037d9434ad5c138eb8067be. Last accessed 11 February 2020.

Collins, S., Pockman, W. & Fargione, J. (2016). Warming-El Nino-Nitrogen Deposition Experiment (WENNDEx): Net primary production quadrat data at the Sevilleta National Wildlife Refuge, New Mexico (2006–present) ver 234294. Environmental Data Initiative. https://doi.org/10.6073/pasta/188fc20d97b750408e764bb9d2e067e9.

Cousins, S.A., Ohlson, H. & Eriksson, O. (2007). Effects of historical and present fragmentation on plant species diversity in semi-natural grasslands in Swedish rural landscapes. *Landsc. Ecol.*, 22, 723–730.

Cusser, S., Bahlai, C., Swinton, S.M., Robertson, G.P. & Haddad, N.M. (2020). Long-term research avoids spurious and misleading trends in sustainability attributes of no-till. *Global Change Biol.*, 26(6), 3715–3725.

Deegan, L. (2012). Nekton species counts and density from flume net collections along rowley river tidal creeks associated with long term fertilization experiments, Rowley, MA. Environmental Data Initiative. https://doi.org/10.6073/pasta/fc88326097d92c1afccce2b3e5e7ab30. Last accessed 11 February 2021.

Deegan, L. & Warren, R. (2012). Marsh plant species stem counts for rowley river tidal creeks associated with long term fertilization experiments, Rowley and Ipswich, MA. Environmental Data Initiative. https://doi.org/10.6073/pasta/87da5b252da19b4c95b883f6206b176b. Last accessed 11 February 2021.

Del-Val, E.K. & Crawley, M.J. (2005). Are grazing increaser species better tolerators than decreasers? An experimental assessment of defoliation tolerance in eight British grassland species. *J. Ecol.*, 93, 1005–1016.

Didham, R.K., Basset, Y., Collins, C.M., Leather, S.R., Littlewood, N.A., Menz, M.H. *et al.* (2020). Interpreting insect declines: seven challenges and a way forward. *Insect Conserv. Divers.*, 13(2), 103–114.

Ellis, C.J. & Coppins, B.J. (2007). 19th Century woodland structure controls stand scale epiphyte diversity in present day Scotland. *Divers. Distrib.*, 13, 84–91.

Elton, C. & Nicholson, M. (1942). The ten-year cycle in numbers of the lynx in Canada. *J. Anim. Ecol.*, 11, 215–244.

Estes, L., Elsen, P.R., Treuer, T., Ahmed, L., Caylor, K., Chang, J. *et al.* (2018). The spatial and temporal domains of modern ecology. *Nat. Ecol. Evol.*, 2(5), 819–826.

Folke, C. (2006). Resilience: The emergence of a perspective for social–ecological systems analyses. *Glob. Environ. Change*, 16, 253–267.

Fournier, A.M., White, E.R. & Heard, S.B. (2019). Site-selection bias and apparent population declines in long-term studies. *Conserv. Biol.*, 33(6), 1370–1379.

Gonzalez, A., Cardinale, B.J., Allington, G.R., Byrnes, J., Arthur Endsley, K., Brown, D.G. *et al.* (2016). Estimating local biodiversity change: a critique of papers claiming no net loss of local diversity. *Ecology*, 97, 1949–1960.

Gonzalez, A. & Chaneton, E.J. (2002). Heterotroph species extinction, abundance and biomass dynamics in an experimentally fragmented microecosystem. *J. Anim. Ecol.*, 71, 594–602.

Gonzalez, A., Germain, R.M., Srivastava, D.S., Filotas, E., Dee, L.E., Gravel, D. *et al.* (2020). Scaling-up biodiversity-ecosystem functioning research. *Ecol. Lett.*, 23(4), 757–776.

Grime, J.P. (1998). Benefits of plant diversity to ecosystems: immediate, filter and founder effects. *J. Ecol.*, 86(6), 902–910.

Gustavsson, E., Lennartsson, T. & Emanuelsson, M. (2007). Land use more than 200 years ago explains current grassland plant diversity in a Swedish agricultural landscape. *Biol. Conserv.*, 138, 47–59.

Haddad, N.M., Brudvig, L.A., Clobert, J., Davies, K.F., Gonzalez, A., Holt, R.D. *et al.* (2015). Habitat fragmentation and its lasting impact on Earth's ecosystems. *Sci. Adv.*, 1, e150005.

Haddad, N.M., Tilman, D. & Knops, J.M. (2002). Long-term oscillations in grassland productivity induced by drought. *Ecol. Lett.*, 5(1), 110–120.

Hamburg, S. (2016). Gastropod abundance at Hubbard Brook Experimental Forest, Watershed 1 and West of Watershed 6, 1997–2006. Environmental Data Initiative. https://doi.org/10.6073/pasta/7ab1863e56f8cc823b3c7f966bc8f43d. Last accessed 11 February 2021.

Hanski, I. & Ovaskainen, O. (2002). Extinction debt at extinction threshold. *Conserv. Biol.*, 16, 666–673.

Hastings, A. (2010). Timescales, dynamics, and ecological understanding. *Ecology*, 91, 3471–3480.

Helm, A., Hanski, I. & Pärtel, M. (2006). Slow response of plant species richness to habitat loss and fragmentation. *Ecol. Lett.*, 9, 72–77.

Hughes, B.B., Beas-Luna, R., Barner, A.K., Brewitt, K., Brumbaugh, D.R., Cerny-Chipman, E.B. *et al.* (2017). Long-term studies contribute disproportionately to ecology and policy. *Bioscience*, 67, 271–281.

Ives, A.R. & Carpenter, S.R. (2007). Stability and diversity of ecosystems. *Science*, 317, 58–62.

Jackson, S.F., Walker, K. & Gaston, K.J. (2009). Relationship between distributions of threatened plants and protected areas in Britain. *Biol. Conserv.*, 142, 1515–1522.

Jarvis, M.F. & Williams, M. (2016). Irreproducibility in preclinical biomedical research: perceptions, uncertainties, and knowledge gaps. *Trends Pharmacol. Sci.*, 37, 290–302.

Joern, A. (2020). CGR02 Sweep sampling of grasshoppers on Konza Prairie LTER watersheds ver 14. Environmental Data Initiative, https://doi.org/10.6073/pasta/7b2259dcb0e499447e0e11dfb562dc2f. Last accessed 11 February 2021.

van Klink, R., Bowler, D.E., Gongalsky, K.B., Swengel, A.B., Gentile, A. & Chase, J.M. (2020). Meta-analysis reveals declines in terrestrial but increases in freshwater insect abundances. *Science*, 368, 417–420.

Knapp, A.K., Smith, M.D., Hobbie, S.E., Collins, S.L., Fahey, T.J., Hansen, G.J.A. *et al.* (2012). Past, present, and future roles of long-term experiments in the LTER Network. *Bioscience*, 62, 377–389.

Kratz, T. (2019). Little rock lake experiment at North Temperate Lakes LTER: Zooplankton count 1983–2000 ver 3. Environmental Data Initiative. https://doi.org/10.6073/pasta/86d37b9e57c56038c492853ec3aa2d0a. Last accessed 11 February 2021.

Krauss, J., Bommarco, R., Guardiola, M., Heikkinen, R.K., Helm, A., Kuussaari, M. *et al.* (2010). Habitat fragmentation causes immediate and time delayed biodiversity loss at different trophic levels. *Ecol. Lett.*, 13, 597–605.

Kuussaari, M., Bommarco, R., Heikkinen, R., Helm, A., Krauss, J., Lindborg, R. & Öckinger, E. (2009). Extinction debt: a challenge for biodiversity conservation. *Trends Ecol. Evol.*, 24, 564–571.

Landis, D. (2018). Insect population dynamics on the main cropping system experiment at the Kellogg Biological Station, Hickory Corners, MI (1989 to 2017). Environmental Data Initiative. https://doi.org/10.6073/pasta/6b6cc0ad7897d9008e8cf918bbf552d2. Last accessed 11 February 2021.

Lawes, J.B. & Gilbert, J.H. (1880). X. Agricultural, botanical, and chemical results of experiments on the mixed herbage of permanent meadow, conducted for more than twenty years in succession on the same land.—Part I. *Philos. Trans. R. Soc. Lond.*, 171, 289–416.

Lawton, J.H. (1994). What do species do in ecosystems? *Oikos*, 71, 367–374.

Lightfoot, D. (2010). Small Mammal Exclosure Study (SMES) vegetation data from the Chihuahuan Desert Grassland and Shrubland at the Sevilleta National Wildlife Refuge, New Mexico (1995–2009). Environmental Data Initiative. https://doi.org/10.6073/pasta/6c758bc7e6e3ea95242d787ee49a263a. Last accessed 11 February 2021.

Lightfoot, D. (2016a). Small Mammal Exclosure Study (SMES) ant data from Chihuahuan Desert Grassland and Shrubland at the Sevilleta National Wildlife Refuge, New Mexico (1995–2005) ver 122263.

Environmental Data Initiative. https://doi.org/10.6073/pasta/c4963aa3363d18ac99bd40307db2249d.

Lightfoot, D. (2016b). Small Mammal Exclosure Study (SMES) rabbit feces data from Chihuahuan Desert Grassland and Shrubland at the Sevilleta National Wildlife Refuge, New Mexico (1995–2005) ver 144612. Environmental Data Initiative. https://doi.org/10.6073/pasta/a98cf83188ecd220320a60b836e92925.

Lightfoot, D. (2016c). Small Mammal Exclosure Study (SMES) vegetation data from the Chihuahuan Desert Grassland and Shrubland at the Sevilleta National Wildlife Refuge, New Mexico (1995–2009) ver 182100. Environmental Data Initiative. https://doi.org/10.6073/pasta/6c758bc7e6e3ea95242d787ee49a263a

Likens, G.E., Driscoll, C.T. & Buso, D.C. (1996). Long-term effects of acid rain: response and recovery of a forest ecosystem. *Science*, 272 (5259), 244–246.

Magurran, A.E., Baillie, S.R., Buckland, S.T., Dick, J.M., Elston, D.A., Scott, E.M. *et al.* (2010). Long-term datasets in biodiversity research and monitoring: assessing change in ecological communities through time. *Trends Ecol. Evol.*, 25(10), 574–582.

Marinos, R.E., Campbell, J.L., Driscoll, C.T., Likens, G.E., McDowell, W.H., Rosi, E.J. *et al.* (2018). Give and take: A watershed acid rain mitigation experiment increases baseflow nitrogen retention but increases stormflow nitrogen export. *Environ. Sci. Technol.*, 52(22), 13155–13165.

McCann, K.S., Cazelles, K., MacDougall, A.S., Fussmann, G.F., Bieg, C., Cristescu, M. *et al.* (2020). Landscape modification and nutrient-driven instability at a distance. *Ecol. Lett.*, https://doi.org/10.1111/ele.13644.

Milchunas, D. (2014). SGS-LTER Effects of grazing on ecosystem structure and function (GZTX): Vegetation basal cover on the Central Plains Experimental Range, Nunn, Colorado, USA 1992–2011, ARS Study Number 32. Environmental Data Initiative. https://doi.org/10.6073/pasta/f8aa046f7bb6533bb2cb52a64cdfb996. Last accessed 11 February 2021.

Plie, L.S., Wang, G.G., Stovall, J.P., Siemann, E., Wheeler, G.S. & Gabler, C.A. (2017). Mechanisms of Chinese tallow (*Triadica sebifera*) invasion and their management implications – a review. *For. Ecol. Manag.*, 404, 1–13.

Pockman, W. (2013). Rainfall manipulation study vegetation data from the Chihuahuan Desert Grassland and Creosote Shrubland at the Sevilleta National Wildlife Refuge, New Mexico (2003–2011) ver 167839. Environmental Data Initiative. https://doi.org/10.6073/pasta/361fbce9ce7d4d9530e34b4a8ee3c02e.

Pounds, J.A., Fogden, M.P. & Campbell, J.H. (1999). Biological response to climate change on a tropical mountain. *Nature*, 398, 611–615.

Rasmussen, P.E., Goulding, K.W.T., Brown, J.R., Grace, P.R., Janzen, H.H. & Korschens, M. (1998). Long-term agroecosystem experiments: Assessing agricultural sustainability and global change. *Science*, 282, 893–896.

Record, S., Voelker, N.M., Zarnetske, P.L., Wisnoski, N.I., Tonkin, J.D., Swan, C. *et al.* (2020). Novel insights to be gained from applying metacommunity theory to long-term, spatially replicated biodiversity data. *Front. Ecol. Evol.*, 8, 479.

Reed, D. & Harrer, S. (2017). SBC LTER: Reef: Long-term experiment: algal biomass. Environmental Data Initiative. https://doi.org/10.6073/pasta/4cc5ac15d0684e1633b27fbfe23c6b8c. Last accessed 11 February 2021.

Rees, M., Condit, R., Crawley, M., Pacala, S. & Tilman, D. (2001). Long-term studies of vegetation dynamics. *Science*, 293(5530), 650–655.

Rosenthal, R., Cooper, H. & Hedges, L. (1994). Parametric measures of effect size. In: *The handbook of research synthesis* (eds Cooper, H., Hedges, L.V. & Valentine, J.C.). Russel Sage Foundation, New York City, pp. 231–244.

Rosi-Marshall, E.J., Bernhardt, E.S., Buso, D.C., Driscoll, C.T. & Likens, G.E. (2016). Acid rain mitigation experiment shifts a forested watershed from a net sink to a net source of nitrogen. *Proc. Natl Acad. Sci.*, 113(27), 7580–7583.

Sagarin, R.D., Gaines, S.D. & Gaylord, B. (2006). Moving beyond assumptions to understand abundance distributions across the ranges of species. *Trends in Ecol. Evol.*, 21, 524–530.

Santa Barbara Coastal LTER and Reed, D. (2020a). SBC LTER: Reef: long-term experiment: kelp removal: invertebrate and algal density ver 17. Environmental Data Initiative. https://doi.org/10.6073/pasta/731d8515e67243716ccb4ee7a28b8843.

Santa Barbara Coastal LTER and Reed, D. (2020b). SBC LTER: Reef: long-term experiment: kelp removal: fish abundance ver 19. Environmental Data Initiative. https://doi.org/10.6073/pasta/ecf2e269db7a4807bcaa765422d8186c.

Sharpe, J. (2017). Ferns surveys of individuals of terrestrial ferns in Canopy Trimming Experiment (CTE) plots document changes in species richness and abundance over time in response to canopy opening and/or debris deposition ver 816303. Environmental Data Initiative. https://doi.org/10.6073/pasta/04ec99ff3b45964420253ef4159a727c. Last accessed 11 February 2021.

Shaver, G. (2016). Above ground biomass in acidic tussock tundra experimental site, 1989, Arctic LTER, Toolik, Alaska. ver 8. Environmental Data Initiative. https://doi.org/10.6073/pasta/668dc98c3dbd83a308f0f38fb833f23e

Smith, M., Collins, S. & Blair, J. (2020). RMP01 Rainfall manipulation plot study at Konza Prairie ver 2. Environmental Data Initiative. https://doi.org/10.6073/pasta/fdd3a94b43a9b67ac8fef6787790a682. Last accessed 11 February 2021.

Staddon, P., Lindo, Z., Crittenden, P.D., Gilbert, F. & Gonzalez, A. (2010). Connectivity, non-random extinction and ecosystem function in experimental metacommunities. *Ecol. Lett.*, 13, 543–552.

Suding, K.N. & Gross, K.L. (2006). The dynamic nature of ecological systems: multiple states and restoration trajectories. In: *Foundations of Restoration Ecology* (eds Falk, D.A., Palmer, M.A. & Zedler, J.B.). Island Press, Washington, DC, pp. 190–209.

Tilman, D. (1982). *Resource competition and community structure*. Princeton University Press, Princeton, NJ.

Tilman, D. (1989). Ecological experimentation: strengths and conceptual problems. In: *Long-term studies in ecology* (ed. Likens, G.E.). Springer, New York, NY, pp. 136–157.

Tilman, D. (1997). Community invasibility, recruitment limitation, and grassland biodiversity. *Ecology*, 78(1), 81–92.

Tilman, D. (2018a). Plant aboveground biomass data: long-term nitrogen deposition: population, community, and ecosystem consequences. Environmental Data Initiative. https://doi.org/10.6073/pasta/93b27926879861815ebb021bfd6f14ae. Last accessed 11 February 2021.

Tilman, D. (2018b). Main plots all arthropod insect sweepnet sampling 1996–200: Biodiversity II: effects of plant biodiversity on population and ecosystem processes. Environmental Data Initiative. https://doi.org/10.6073/pasta/4c1795e6769bf78e3c947e92db75eef6. Last accessed 11 February 2021.

Tilman, D. (2018c). Plant aboveground biomass data: nitrogen addition and dynamics of recovery from cessation of N addition. Environmental Data Initiative. https://doi.org/10.6073/pasta/ef24de9bc55442766123fc02a405c13f. Last accessed 11 February 2021.

Tilman, D., May, R.M., Lehman, C.L. & Nowak, M.A. (1994). Habitat destruction and the extinction debt. *Nature*, 371(6492), 65–66.

Voelkl, B. & Würbel, H. (2016). Reproducibility crisis: are we ignoring reaction norms. *Trends Pharmacol. Sci.*, 37, 509–510.

Wall, D. & Virginia, R. (2016). McMurdo dry valleys soil long term manipulation experiment ver 8. Environmental Data Initiative. https://doi.org/10.6073/pasta/4d895dea5dc6f4650a0da3078a3febc5. Last accessed 11 February 2021.

Watts, K., Whytock, R.C., Park, K.J., Funtes-Montemayor, E., Macgregor, N.A., Duffield, S. & McGowan, P.J.K. (2020). Ecological time lags and the journey towards conservation success. *Nat. Ecol. Evol.*, 4, 304–311.

Wauchope, H.S., Amano, T., Sutherland, W.J. & Johnston, A. (2019). When can we trust population trends? A method for quantifying the effects of sampling interval and duration. *Methods Ecol. Evol.*, 10(12), 2067–2078.

Editor, Jonathan Chase
Manuscript received 12 November 2020
First decision made 1 February 2021
Manuscript accepted 2 February 2021