How do climate change experiments alter plot-scale climate?

A.K. Ettinger[1,2,a], I. Chuine[3,b], B.I. Cook[4,5,c], J.S. Dukes[6,d], A.M. Ellison[7,e], M.R. Johnston[8,f], A.M. Panetta[9,g], C.R. Rollinson[10,h], Y. Vitasse[11,12,i], and E.M. Wolkovich[1,8,13,j]

[1]Arnold Arboretum of Harvard University, Boston, Massachusetts 02131, USA

[2]Tufts University, Medford, Massachusetts 02155, USA

[3]CEFE UMR 5175, CNRS, Université de Montpellier, Université Paul-Valéry Montpellier, EPHE IRD, Montpellier, France

[4]Lamont-Doherty Earth Observatory, Columbia University, Palisades, New York 10964, USA

[5]NASA Goddard Institute for Space Studies, New York, New York 10025, USA

[6]Department of Forestry and Natural Resources and Department of Biological Sciences, Purdue University, West Lafayette, Indiana 47907, USA

[7]Harvard Forest, Harvard University, Petersham, Massachusetts 01366, USA

[8]Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, Massachusetts 02138, USA

[9]Department of Ecology and Evolutionary Biology, University of Colorado, Boulder, Colorado 80309, USA

[10]Center for Tree Science, The Morton Arboretum, Lisle, Illinois 60532, USA

[11]Institute of Geography, University of Neuchâtel, Neuchâtel, Switzerland

[12]Swiss Federal Institute for Forest, Snow and Landscape Research WSL, Neuchâtel, Switzerland

[13]Forest & Conservation Sciences, Faculty of Forestry, University of British Columbia, Vancouver, BC, Canada

[a]Corresponding author; email: aettinger@fas.harvard.edu; phone: 781-296-4821; mailing address: 1300 Centre Street, Boston, Massachusetts 02140, USA

[b]isabelle.chuine@cefe.cnrs.fr

[c]bc9z@ldeo.columbia.edu

[d]jsdukes@purdue.edu

[e]aellison@fas.harvard.edu

[f]mjohnston@g.harvard.edu

[g]anne.panetta@colorado.edu

[h]crollinson@mortonarb.org

[i]yann.vitasse@wsl.ch

[j]e.wolkovich@ubc.ca

July 16, 2018

Statement of authorship All authors conceived of this manuscript, which was inspired by our discussions at a Radcliffe Exploratory Seminar in 2016, and all authors contributed to manuscript revisions. AKE and EMW conceived of the idea for the literature review, database compilation, and related Radcliffe Exploratory Seminar. AKE compiled the datasets; AKE and CRR analyzed the data and created the figures; AKE wrote the manuscript.

Data Accessibility The MC3E database will be available at KNB (Ettinger & Wolkovich, 2018), along with all R code from the analyses included in this paper. (Currently, metadata are published there; the full database and R code are available to reviewers on github.)

Running title Experimental climate change

Key words global warming, warming experiment, microclimate, soil moisture, spring phenology, budburst, direct and indirect effects, structural control, hidden treatment, active-warming, target temperature, feedback

Type of article Review and Synthesis

Number of words in abstract 200

Number of words in main text 4,933

Number of references 97

Number of figures 5

Number of tables 0

Number of text boxes 1

Number of words in Box 1 570

Abstract

To understand and forecast biological responses to climate change, scientists frequently use field experiments that alter temperature and precipitation. Climate manipulations can manifest in complex ways, however, challenging interpretations of biological responses. We reviewed publications from active-warming experiments to compile a database of daily plot-scale climate data from 15 experiments that use forced air, infrared heaters, or soil cables to warm plots. We find that the common practices of analyzing primarily mean changes among treatments and analyzing treatments as categorical variables (e.g., warmed verses unwarmed) masks important variation in treatment effects over space and time. Our synthesis showed that measured mean warming, in plots with the same target warming, varied by 1.6°C, on average, among blocks within a study. Furthermore, warming treatments produce non-temperature effects as well, such as soil drying. The implications of these complexities can have important biological consequences. We show one such consequence with a case study of plant phenology, in which accounting for drier soils with warming triples the estimated sensitivity of budburst to temperature. Based on our synthesis, we present recommendations for future analyses, experimental design, and data sharing that will improve the ability of climate change experiments to accurately identify and forecast species' responses.

Introduction

Climate change is dramatically altering earth's biota, shifting the physiology, distribution, and abundance of organisms, with cascading community, ecosystem, and climate effects (Shukla & Mintz, 1982; Cox *et al.*, 2000; Thomas *et al.*, 2004; Parmesan, 2006; Field *et al.*, 2007; Sheldon *et al.*, 2011; Urban *et al.*, 2012). Much uncertainty exists about how particular individuals, populations, species, communities, and ecosystems will respond as warming becomes more extreme (Thuiller, 2004; Friedlingstein *et al.*, 2014). Predicting biological responses to current and future climate change—and their feedbacks to earth's climate and ecosystem services—is one of the most significant challenges facing ecologists today.

Two common approaches for understanding biological effects of climate change are observational studies, which correlate recorded biological patterns with measured trends in climate, and process-based modeling; yet these approaches are insufficient for several reasons. Observational studies and correlative models cannot disentangle the causal effects of warming (one aspect of climate) from other factors that have also changed over time, such as successional stage or land use. In addition, models based on correlative data may fail to make useful predictions for future conditions that fall outside the range of historical variability (e.g., Pearson & Dawson, 2004; Hampe, 2004; Ibanez *et al.*, 2006; Swab *et al.*, 2012; Chuine *et al.*, 2016). Climate change will yield warmer temperatures than the previous 150 years, and possibly warmer than at any time in the last 2000 years (Ohlemüller *et al.*, 2006; Williams & Jackson, 2007; Williams *et al.*, 2007; Stocker *et al.*, 2013). Process-based models overcome some of these challenges through inclusion of explicit mechanistic relationships between climate and biological outcomes. However, they are limited by the processes they include (i.e., our understanding of mechanism), as well as by the data available to parameterize those processes (Moorcroft, 2006; Kearney & Porter, 2009).

Experimental data from field-based climate change experiments are crucial to fill these knowledge gaps and determine mechanistic links between climate change and biological responses. Experiments can quantify biological responses to different levels of climate change, and can create the "no-analog" climate scenarios forecasted for the future, particularly when they employ active-warming methods, such as forced air heaters, soil warming cables, or infrared heaters (Shaver *et al.*, 2000; Williams *et al.*, 2007; Aronson & McNulty, 2009). In addition, active-warming can be combined with precipitation manipulations (e.g., snow removal, water additions or reductions), offering the ability to assess individual and interactive effects of temperature and precipitation, separate from other environmental changes (e.g., Price & Waser, 1998; Cleland *et al.*, 2006; Sherry *et al.*, 2007; Rollinson & Kaye, 2012). Compared with indoor growth-chamber experiments, field-based experiments offer the possibility of preserving important but unknown or unquantified feedbacks among biotic and abiotic components of the studied systems.

With climate change experiments, ecologists often aim to test hypotheses about how projected warming will affect species' growth, survival, and future distributions (Dukes & Mooney, 1999; Hobbie *et al.*, 1999; Morin *et al.*, 2010; Pelini *et al.*, 2011; Chuine *et al.*, 2012; Reich *et al.*, 2015; Gruner *et al.*, 2017). Recent research suggests, however, that climate manipulations may not always alter plot-scale climate (hereafter, microclimate) in ways that are consistent with observed changes over time (Wolkovich *et al.*, 2012; Menke *et al.*, 2014; Polgar *et al.*, 2014; Andresen *et al.*, 2016). For extrapolation of experimental findings to the real world, we need detailed assessments of how active-warming experiments alter the microclimate conditions experienced by organisms, and the extent to which these conditions are similar to current field conditions or anticipated climate change.

Here, we investigate the complex ways that microclimate is altered by active-warming treatments, both directly and indirectly, across multiple studies. The qualitative challenges and opportunities of climate change experiments have been summarized previously (e.g., De Boeck *et al.*, 2015) and effects of these manipulations on some aspects of microclimate have been published for individual sites (e.g., Harte *et al.*, 1995; McDaniel *et al.*, 2014b; Pelini *et al.*, 2011). However, our quantitative meta-analysis allows us to examine trends across sites and warming designs, and make recommendations based on this information. Using plot-level daily microclimate data from 15 active-warming experiments (yielding 59 experiment years and 14,913 experiment days; Table S1), we show the direct and indirect ways that experimental manipulations alter microclimate. We use a case study of spring plant phenology to demonstrate how analyses that assume a constant warming effect and do not include non-temperature effects of warming treatments on biological responses lead to inaccurate quantification of plant sensitivity to temperature shifts. Finally, we synthesize our findings to make recommendations for future analysis and design of climate change experiments (Box 1).

MicroClimate from Climate Change Experiments (MC3E) database

To investigate how climate change experiments alter microclimate, we first identified published, active-warming field experiments, many of which included precipitation manipulations. We focused on *in situ* active-warming manipulations because recent analyses indicate that active-warming methods are the most controlled and consistent methods available for experimental warming (Kimball, 2005; Kimball *et al.*, 2008; Aronson & McNulty, 2009; Wolkovich *et al.*, 2012). We do not include passive-warming experiments because they have been analyzed extensively already and are known to have distinct issues, including reduction in wind, overheating, and great variation in the amount of warming depending on irradiance and snow depth (Marion *et al.*, 1997; Shaver *et al.*, 2000; Wolkovich *et al.*, 2012; Bokhorst *et al.*, 2013, see also Table S2).

We carried out a full literature review to identify potential active-warming field experiments to include in the database. We followed the methods and search terms of Wolkovich *et al.* (2012) for their Synthesis of Timings Observed in iNcrease Experiments (STONE) database (Wolkovich *et al.*, 2012), but restricted our focus to active-warming experiments. Further, because our goal was to tease out variation in microclimate (including temperature and soil moisture), we focused on warming studies that included both/either multiple levels of warming and/or precipitation treatments. These additional restrictions constrained the list to 11 new studies published after the STONE database, as well as six of the 37 studies in the STONE database.

We contacted authors to obtain daily microclimate and phenological data for these 17 studies and received data (or obtained publicly available data) for 10 of them, as well as datasets from five additional sites offered or suggested to us over the course of our literature review and data analysis. The daily temperature and soil moisture data from these 15 experiments comprise the MicroClimate from Climate Change Experiments (MC3E) database (Figures 1 and S1, Table S1), which is available at KNB (Ettinger & Wolkovich, 2018). We examined how these experiments altered microclimate, using mixed-effects models that allowed for inherent differences among studies (through a random effect of study on the intercept), while also estimating various across-study effects, such as degree of warming or warming type.

Complexities in interpreting experimental climate change

Climate change experiments often include detailed monitoring of climate variables at the plot-level, yielding large amounts of data, such as daily or hourly temperature and other climate variables, over the course of an experiment. Ecologists, however, are generally interested in the ecological responses (e.g., community dynamics, species' growth, abundance, or phenology), which are collected on much coarser timescales (e.g., weekly or annually). Not surprisingly, then, when analyzing ecological responses, authors typically provide detailed information on the observed biological responses, and report only the mean change in climate over the course of the experiment and whether it matched their target level of change (e.g., Price & Waser, 1998; Rollinson & Kaye, 2012; Clark *et al.*, 2014a,b). Several studies have conducted detailed, independent analyses of microclimate data from warming experiments (e.g., Harte *et al.*, 1995; Kimball, 2005; Kimball *et al.*, 2008; McDaniel *et al.*, 2014b; Pelini *et al.*, 2011). While these detailed analyses provide valuable case studies of experimental effects on microclimate data alone, they have generally not been incorporated into analyses of ecological responses.

In interpreting ecological responses to climate change manipulations, the focus has been primarily on mean shifts in microclimate, but the imposed manipulations result in much more complex shifts. The magnitude of change in these manipulations varies in time and space, and the presence of experimental equipment alone (with no heat added) often alters environmental conditions. These factors, discussed below, challenge our interpretation of how experimental warming studies forecast effects of climate change on organisms and ecosystems.

Effects on microclimate vary over time and space

Reporting only the mean temperature difference across the duration of a warming study masks potentially important temporal variation in temperature among treatments (compare Figure 2 to Figure S2). Using the MC3E database, we found that active-warming reduces the range of above-ground daily temperature by 0.37°C per °C of target warming (Table S3, see also Table S1, which details the different methods used to measure and warm temperatures). Active-warming decreased above-ground daily temperature range by differentially affecting maximum and minimum temperatures: warming increased daily minima by 0.81°C per °C of target warming, but only increased daily maxima by 0.48°C per °C of target warming (Table S3). These effects varied by site (Table S3), but we found no clear patterns by warming type (e.g., infrared versus forced air). Soil daily temperature range was not affected by experimental warming, as warming altered minimum and maximum daily temperatures similarly (Table S4).

We observed strong seasonal and annual variations in the effects of experimental warming (Figures 1, 2, Table S5). Warming generally appears close to targets in winter and early spring, and farthest below targets in summer (day of year 150-200, when evapotranspiration within a robust plant canopy can dissipate energy and act to cool vegetation surfaces), though patterns differ among sites (Figure 1). The variation in warming effectiveness may be driven by interactions between warming treatments and daily, seasonal, and annual weather patterns, since the magnitude of warming can vary as weather conditions change. Both infrared heaters and soil cables fail to achieve target temperature increases during rainstorms (Peterjohn *et al.*, 1993; Hoeppner & Dukes, 2012) and with windy conditions (Kimball, 2005; Kimball *et al.*, 2008). In addition, treatments are often applied inconsistently within or across years. Heat applications are frequently shut off during winter months, and some heating methods, even if left on throughout the year, do not warm consistently (e.g., Clark *et al.*, 2014a,b; Hagedorn *et al.*, 2010).

Treatment effects also vary spatially, further complicating interpretation of climate change experiments. The MC3E database contains six studies that used blocked designs, allowing us to examine spatial variation in the amount of warming (i.e., the difference between treatment and control plots within a block). These studies include five infrared and one soil warming cable experiment. We found that the amount of observed warming frequently varied by more than 1°C (mean= 1.6°C, maximum = 3°C) among blocks (Figure 2, Table S6); this variation in warming is substantial, as it is equivalent to the target warming treatment for many studies. The differences in warming among blocks may be caused by fine-scale variation in vegetation, slope, aspect, soil type, or other factors that can alter wind or soil moisture, which in turn affect warming (Peterjohn *et al.*, 1993; Kimball, 2005; Kimball *et al.*, 2008; Hoeppner & Dukes, 2012; Rollinson & Kaye, 2015).

Of course, identical experimental treatments across space and time are neither necessary, nor realistic, for robust analysis of experimental results and forecasting. Indeed, the spatial and temporal variation we report could improve and refine models, and—at least in some regions—may be consistent with contemporary patterns of climate change (Stocker *et al.*, 2013). Taking advantage of this variation, though, requires understanding and reporting it (e.g., Milcu *et al.*, 2016). However, because fine-scale and temporal variations in warming treatments are rarely analyzed explicitly with ecological data, the implications for interpretation of experimental findings are unclear.

Experimental infrastructure alters microclimate

Experimental structures themselves can alter temperature and other important biotic and abiotic variables in ways that are not generally examined in experimental climate change studies. The importance of controls that mimic a treatment procedure without actually applying the treatment is widely acknowledged in biology (e.g., Dayton, 1971; Spector, 2001; Johnson & Besselsen, 2002; Quinn & Keough, 2002). Though some experimental climate change studies include treatments with non-functional warming equipment as well as ambient controls, the magnitude and effects of experimental infrastructure alone on climate are rarely interpreted or analyzed.

To investigate the magnitude of infrastructure effects, we compared temperature and soil moisture data from five active-warming studies at two sites: Duke Forest and Harvard Forest (Farnsworth *et al.*, 1995; Clark *et al.*, 2014b; Marchin *et al.*, 2015; Pelini *et al.*, 2011)(see Supplemental Materials for model details). These were the only studies in the MC3E database that monitored climate in two types of control plots: structural controls (i.e., 'shams' or 'disturbance controls,' which contained all the warming infrastructure, such as soil cables (n=1), forced air chambers (n=2), or both (n=2), but with no heat applied) and ambient controls with no infrastructure added. Other studies monitored environmental conditions in only structural controls (n=5) or ambient controls (n=4). We were unable to compare ambient and structural controls for experiments using infrared heating, because no studies in our database included both control types. (A separate analysis suggested that there may be infrastructure effects on microclimate for infrared studies in our database; see Supplemental Materials, especially Table S7).

We found that experimental structures altered above-ground and soil temperatures in opposing ways: above-ground temperatures were higher in the structural controls than in ambient controls, whereas soil temperatures were lower in structural controls compared with ambient controls (Figure 3a-d). This general pattern was consistent across different temperature models (mean, minimum, and maximum temperatures), although the magnitude varied among seasons, studies, and years (Figure 3a-d, Tables S8-S11). We also found that experimental infrastructure decreased soil moisture relative to ambient conditions across all seasons, studies, and years (Figure 3e, Tables S12, S13).

There are several possible reasons for the observed climatic differences between ambient and structural controls. Infrastructure materials may shade the plots, reduce airflow, reduce albedo relative to surroundings, or otherwise change the energy balance. Specifically, soil temperatures may be cooler in structural controls for forced air studies because the experimental structures block sunlight from hitting the ground surface, causing less radiative heating of the ground in structural controls compared to ambient controls. In addition, above-ground temperatures may be warmer in structural controls because the structures radiatively warm the air around them and block wind, inhibiting mixing with air outside of the plot. Structures may also interfere with precipitation hitting the ground, thereby reducing local soil moisture and snowpack, with its insulative properties. Finally, for some warming types (e.g., soil cables), structural controls experience increased soil disturbance compared with ambient controls; this may alter water flow and percolation, and introduce conductive material via the cables or posts.

To the extent that differences between ambient and structural controls have been reported in previous studies, our findings appear to be consistent. Clark *et al.* (2014b), who used forced air and soil cables for warming, state that "control of the air temperature was less precise, in part due to air scooping on windy days." Marchin *et al.* (2015), who used forced air warming, note that structural controls had mean spring air temperatures about 0.5°C or more above ambient temperatures. Peterjohn *et al.* (1994), who warmed soil with heating cables, reported cooler soil temperatures in structural controls than in ambient controls at shallow soil depths. Similarly, we found the greatest difference in soil temperature between structural and ambient controls in shallow soils (e.g., exp10, soil depth = 2cm). If addressed, the focus to date has been largely on these abiotic impacts of experimental structures, but structures may also alter herbivory and other biotic conditions (Kennedy, 1995; Moise & Henry, 2010; Wolkovich *et al.*, 2012; Hoeppner & Dukes, 2012).

Our analyses suggest that warming experiments that calculate focal response variables relative to ambient controls (e.g., Price & Waser, 1998; Dunne *et al.*, 2003; Cleland *et al.*, 2006; Morin *et al.*, 2010; Marchin *et al.*,

2015) may not adequately account for the ways in which infrastructure affects microclimate. Results from studies reporting only structural controls (e.g., Sherry *et al.*, 2007; Hoeppner & Dukes, 2012; Rollinson & Kaye, 2012), should be cautiously applied outside of an experimental context, as—without ambient controls—their inference is technically limited to the environment of the structural controls. Our results suggest that studies aiming to predict or forecast the effects of climate change on organisms and ecosystems would benefit from employing both structural and ambient controls so that they may separate artifacts due to infrastructure from the effects of experimental warming.

Indirect and feedback effects of climate change manipulations

Climate change experiments often seek to manipulate temperature or precipitation separately as well as interactively, but manipulating either of these variables in isolation is notoriously difficult. Treatments involving precipitation additions typically reduce temperatures in climate change manipulations (Sherry *et al.*, 2007; Rollinson & Kaye, 2012; McDaniel *et al.*, 2014b). For example, Sherry *et al.* (2007) observed that a doubling of precipitation reduced mean air temperatures by 0.44°C, on average, during their one-year observation period.

In the MC3E database, there are four experiments that manipulated both temperature and precipitation, and provided daily above-ground temperature data (three of these also measured soil temperature). Across these studies, all of which used infrared heating, we found that increasing the amount of added precipitation reduced daily minimum and maximum above-ground temperatures, at rates of 0.01 and 0.02°C, respectively, and soil temperatures, at a rate of 0.01°C for both minimum and maximum temperature, per percent increase in added precipitation (Table S14). Thus, a 50% increase in precipitation would be expected to decrease temperature by 0.5°C. This is likely because increasing soil moisture (an effect of precipitation additions) typically shifts the surface energy balance to favor latent (i.e., evapotranspiration) over sensible energy fluxes, reducing heating of the air overlying the soils. Maintaining target warming levels is a challenge even for independent feedback systems, which vary energy inputs using ongoing temperature measurements, particularly during seasons or years with wetter soils and higher evapotranspiration (Rich *et al.*, 2015).

In addition to its effects on temperature, experimental warming often increases vapor pressure deficit and reduces soil water content (e.g., Harte *et al.*, 1995; Sherry *et al.*, 2007; Morin *et al.*, 2010; Pelini *et al.*, 2014;

Templer *et al.*, 2016). Of the 15 experiments in the MC3E database, we examined the 12 that continuously measured and reported soil moisture. We included target warming, warming type, and their interaction as predictors (excluding data from plots with precipitation treatments) and accounted for other differences among studies by including a random effect of study (see Supplemental Materials for details). We found that experimental warming reduced soil moisture across all warming types, with substantial variation among experiments (Figure 5, Table S15). The drying effect varied by warming type (-0.80% for infrared versus -0.33% for forced air, per °C of target warming, Table S16). Soil moisture can be difficult to measure, with high spatial and temporal variation (Famiglietti *et al.*, 1999; Teuling & Troch, 2005), but these results highlight that changes in soil moisture often accompany temperature changes in active-warming experiments.

Warming and precipitation treatments, and their indirect effects on soil moisture and other abiotic factors, can also alter the biotic environment, which may produce cascading effects. Many studies have found shifts from herbaceous to woody plant communities over time with experimental warming (e.g., Rollinson & Kaye, 2012; McDaniel *et al.*, 2014b,a; Harte *et al.*, 2015). These community shifts may affect resource levels, such as moisture, carbon, and nutrient levels in the soil (McDaniel *et al.*, 2014b,a; Harte *et al.*, 2015) and feed back to affect microclimate (Harte *et al.*, 2015).

The presence of these feedback effects is both a strength of and a challenge for climate change experiments. They may represent important and ecologically realistic effects that became apparent only with the *in situ* field experiment. Alternatively, they may represent artifacts that are unlikely to occur outside of an experimental context. Quantifying, interpreting, and reporting these non-temperature effects in experiments is critical to distinguish these possibilities and to understand mechanisms underlying observed biological responses to climate change.

The widespread presence of indirect effects of climate manipulations highlights the importance of measuring environmental conditions at the plot-level, and using these measurements in analysis and interpretation of results. Many papers published on climate change experiments—including 10 of the 15 references listed in Table S1—analyze warming and/or precipitation treatments as simple categorical predictors (e.g., as in a two-way ANOVA). Our findings, however, demonstrate a clear need for alternative modelling approaches to fully understand the experimental results and to make mechanistic links between changes in climate and ecological responses. One straightforward alternative is to include the continuous climate data (e.g., plot-level temperatures) as predictors of the focal response variable, such as phenological state or species density (e.g., Marchin *et al.*, 2015; Pelini *et al.*, 2014).

Ecological implications

We have highlighted a suite of factors that complicate interpretation of climate change experiments. These indirect effects are similar to the "hidden treatments" described by Huston (1997) in biodiversity experiments, and are likely to have biological implications for many of the responses studied in warming experiments (e.g., Figure 5). Interpretation of experimental climate change effects on biological responses may be misleading because the intended climate treatments (i.e., categorical comparisons or target warming levels) are often used as explanatory variables in analyses (Table S1). The interpretation is likely to be altered by using fine-scale, measured climate as explanatory variables. For example, biological responses may be muted (Figure 5b) or exaggerated (Figure 5c) when direct and indirect effects of climate manipulations interact.

To investigate the ecological implications of non-target abiotic responses to climate warming, we used a simple case study of plant phenology. We used the MC3E database to test if estimates of the temperature sensitivity of phenology vary when calculated using target warming versus plot-level climate variables. We fit two separate mixed-effects models, that differed in their explanatory variables: one used target warming and one used measured climate. Both models had budburst day of year as the response variable, and both included random effects of study (which modeled other differences between studies, that may have affected phenology), year (nested within study, which modeled differences due to weather variability among years that may have altered phenology), and species (which often vary in their phenology). All random effects were modeled on the intercept only; see Supplemental Materials for details.

We found that phenological temperature sensitivity estimates from the two modeling approaches varied three-fold. The target warming model estimated temperature sensitivity of budburst to be -1.91 days/°C (95% CI -2.17, -1.86; Table S17, solid black line in Figure 6), whereas the measured climate model estimated temperature sensitivity of budburst to be -6.00 days/°C (95% CI: -6.74, -5.26; Table S17). Further, all measured climate models with both temperature and moisture had improved model fit compared to the target warming model (Table S18). The best-fit model included mean daily minimum above-ground temperature, mean winter soil moisture, and their interaction as explanatory variables, suggesting that these variables are important drivers of budburst timing (Tables S17, S18). In addition, the measured climate model estimated a significant effect of soil moisture on budburst of -1.51 days/% VWC (95% CI: -1.76, -1.26; Table S17, Figure 6). This negative effect is expected, if reducing moisture delays budburst (Table S17, Figure 6), and is consistent with previous work showing that budburst requires water uptake (Essiamah & Eschrich, 1986).

The increase in estimated temperature sensitivity with measured (rather than target) temperature has two major causes. First, plot-level warming often does not reach target levels (Figure 2), producing a muted effect of temperature in models using target warming. Second, experimental warming's dual effects of decreasing soil moisture and increasing temperature impact budburst in contrasting ways. Decreasing soil moisture has a delaying effect on budburst phenology, opposing the advancing effect of rising temperatures (Figure 5b); thus the effect of temperature is underestimated when moisture is not included in the model. This example shows how the common method of using target warming alone, or even measured temperature alone as done in previous analyses of the particular experiments included here (exp01, exp03, exp04, exp10, Clark *et al.*, 2014a,b; Polgar *et al.*, 2014; Marchin *et al.*, 2015), to understand biological responses may yield inaccurate estimates of temperature sensitivity in warming experiments. In this case, the underestimation may be substantial enough to account for previously described discrepancies between phenological responses to warming in observational versus experimental studies (Wolkovich *et al.*, 2012; Polgar *et al.*, 2014), though further investigation is required.

Accounting for both direct and indirect effects of warming is critical for accurate interpretation of the consequences of climate change (Kharouba *et al.*, 2015). Of particular importance is the extent to which abiotic and biotic effects are realistic forecasts of future shifts that are likely to occur with climate change, or due to artifacts that are unlikely to occur outside of experimental systems (Hurlbert, 1984; Moise & Henry, 2010; Diamond *et al.*, 2013). For many important climatic and ecological metrics, experimental findings of abiotic and biotic effects appear to be consistent with observations. Altered above-ground daily temperature range (i.e., temperature minima changing more than maxima, Table S3) with experimental warming is consistent with observed changes in many places, at least for some time periods. Global minimum temperatures increased more rapidly than maximum temperatures from 1950-1980, reducing above-ground daily temperature range (Thorne *et al.*, 2016; Vose *et al.*, 2005). In addition, the acclimation response of leaf respiration to temperature (Aspinwall *et al.*, 2016; Reich *et al.*, 2016), responses of soil respiration to warming (Carey *et al.*, 2016), and declines in soil carbon at one site (Harte *et al.*, 2015), also appear to be consistent across experiments and observations. These cases suggest that many responses observed in climate change experiments may be accurate harbingers of future biological responses to climate change.

In contrast, some responses documented in climate change experiments may not be in line with future climate change—or may be too uncertain for robust prediction, and thus need explicit analyses and cautious interpretation. For example, soil drying in conjunction with future warming is forecasted in some regions, such as the southwestern United States, mainly because of reductions in precipitation and increased evaporative demand associated with warmer air (Dai, 2013; Seager *et al.*, 2013). The northeastern United States, on the other hand, has been trending wetter over time (Shuman & Burrell, 2017), even though temperatures have warmed. Future changes in soil moisture are uncertain, and likely to vary by region, season, and even soil depth (Seager *et al.*, 2014; Berg *et al.*, 2017). Thus, researchers should not assume that the soil drying observed in warming experiments is likely to occur at all sites with future warming. The uncertainty associated with forecasting changes to soil moisture makes replicating future water availability regimes in climate change experiments especially challenging; one way to meet this challenge and make predictions—even given high uncertainty—is to estimate soil moisture effects in climate change experiments. The altered light, wind, and herbivory patterns documented under experimental infrastructure (Kennedy, 1995; Moise & Henry, 2010; Wolkovich *et al.*, 2012; Hoeppner & Dukes, 2012; Clark *et al.*, 2014b) represent other non-temperature effects that may be potential experimental artifacts and are worth quantifying in future analyses to provide improved estimates of temperature sensitivity.

An additional challenge in relating experiments to observations is that experimental findings may not scale up in space and time. Short-term responses to climate change frequently differ from long-term responses (Woodward, 1992; Elmendorf *et al.*, 2012; Andresen *et al.*, 2016; Reich *et al.*, 2018). Differences may be, in part, because many experiments typically impose some mean shift in climate, but patterns of climate change are likely to be more variable. Many climate models project complex shifts in precipitation: more intense extreme precipitation events (e.g., heavy downpours), more dry days (i.e., less total precipitation events), or both (Polade *et al.*, 2014). In addition, the small spatial scale of experiments may result in responses that are unlikely to be observed at larger scales (Woodward, 1992; Menke *et al.*, 2014). Experimental plots range in area from 1.5 to 36 square meters (Table S1), which may be too small to encapsulate, for example, the rooting zones of perennial plants (Canadell *et al.*, 1996), or foraging ranges for animals (Menke *et al.*, 2014). One approach to overcome these challenges is to conduct larger, longer experiments (Woodward, 1992), though this frequently is not logistically possible and does not easily address how to capture potential shifts in climate variability.

Conclusions

As climate change continues across the globe, ecologists are challenged to not only document impacts, but also make quantitative, robust predictions. Our ability to meet this challenge requires a nuanced mechanistic understanding of how climate directly and indirectly alters biological processes. Climate change experiments, which have been underway for nearly four decades (e.g., Tamaki *et al.*, 1981; Carlson & Bazzaz, 1982; Melillo *et al.*, 2017), provide invaluable information about biological responses to climate change. Yet the full range of changes in environmental conditions imposed by these experiments is rarely presented. We have compiled the first database of microclimate data from multiple warming experiments and shown how time, space, experimental artifacts, and indirect effects of treatments may complicate simple interpretations of these experimental results. We hope this work provides a foundation for gaining the most knowledge and utility from existing experiments via robust analyses, for designing new experiments (see Box 1), and for improved understanding of biological responses to a changing world.

Acknowledgements

We are grateful to those who shared their experimental climate data with us, allowing it to be included in the MC3E database. We thank the Radcliffe Institute for Advanced Study at Harvard University, which provided funding for an Exploratory Seminar at which the ideas in this paper were conceived, and we thank three anonymous reviewers. This research was also supported by the National Science Foundation (NSF DBI 14-01854 to A.K.E.). Any opinion, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Box 1: Recommendations for future climate change experiments

1. *Collect and analyze plot-level climate data.* This includes analyzing and interpreting minimum and maximum values, as well as variance and critical thresholds (e.g., the number and duration of freeze-thaw events and accumulated chilling hours, McDaniel *et al.*, 2014b; Vasseur *et al.*, 2014). We suggest saving the raw data from data loggers (often collected at hourly or higher resolution) to allow quantification of variance (and other summaries) at different temporal resolutions. In assessing which frequency of

measurements is most appropriate for analyses (e.g., hourly, twice daily), it is critical to consider the chronobiology of the event and organisms of interest. For ants, this might mean that temperatures be monitored every minute (Helm & Shavit, 2017); for bacteria, even more frequently.

2. *Analyze measured climate variables rather than targets.* There can be substantial variation in the effects of warming and precipitation treatments among plots and across time (Figure 2). Analyzing measured climate will allow much more in-depth understanding of the drivers and biological effects of variation in temperature and moisture.

3. *Publish high quality, usable data and metadata.* Given that climate manipulations are logistically challenging and expensive (Aronson & McNulty, 2009), and that they often produce a large volume of fine-scale climate data, good curation and data sharing will ensure wider use and deeper understanding of these valuable data. When studying biological implications of a global challenge as large as climate change, progress will come from designing and reporting experiments in ways that facilitate an eventual global data set. Researchers should also be explicit in their warming design (e.g., infrared heating with feedback control or forced air heating with constant energy inputs) to aid future analyses of the performance of different designs, across sites and over time.

4. *Include both structural and ambient controls* and collect, use, and report microclimate and biological data within them. Fewer than half of the studies in our MC3E database reported microclimate data from these two control types (6 out of 15 studies); however, all experiments that did include both control types showed significant effects of infrastructure (Figure 3).

5. *Design relevant manipulations* by consulting observational records and forecasts, including seasonal and annual variation in projected warming. When it is not possible or desirable to match anticipated changes in climate, studies should report how imposed treatments compare to projected changes and past observations (e.g., Hoover *et al.*, 2014; Zhu *et al.*, 2016). In addition, if continuous treatments are not applied throughout the study, we recommend reporting the seasonality and timing of treatments and monitoring the climate throughout the year.

6. *Maximize the duration of climate change experiments* by running some experiments for as long as possible, since the magnitude of climate change treatments can vary considerably among years (Figure 2). In addition, long-term responses of individuals and populations can differ from transient responses (Saleska *et al.*, 2002; Franklin, 1989; Giasson *et al.*, 2013; Harte *et al.*, 2015). We were able to acquire

data extending for ≥5 years for only one study in the MC3E database (exp01), restricting our ability to investigate the effect of study length on experimental climate change.

7. *Conduct syntheses across studies.* As more detailed data are published from experimental climate change studies in divergent ecosystems and warming types, meta-analyses will advance our understanding of the ways that warming affects microclimate and biotic interactions. For example, it would be useful to compare microclimate data among studies using infrared warming applied with constant energy inputs versus infrared warming that varies energy inputs based on measured temperatures.

References

Andresen, L.C., Müller, C., de Dato, G., Dukes, J.S., Emmett, B.A., Estiarte, M., Jentsch, A., Kröel-Dulay, G., Lüscher, A., Niu, S. *et al.* (2016). Shifting impacts of climate change: long-term patterns of plant response to elevated co2, drought, and warming across ecosystems. In: *Advances in ecological research*. Elsevier, vol. 55, pp. 437–473.

Aronson, E.L. & McNulty, S.G. (2009). Appropriate experimental ecosystem warming methods by ecosystem, objective, and practicality. *Agricultural and Forest Meteorology*, 149, 1791–1799.

Aspinwall, M.J., Drake, J.E., Campany, C., Vårhammar, A., Ghannoum, O., Tissue, D.T., Reich, P.B. & Tjoelker, M.G. (2016). Convergent acclimation of leaf photosynthesis and respiration to prevailing ambient temperatures under current and warmer climates in eucalyptus tereticornis. *New Phytologist*, 212, 354–367.

Berg, A., Sheffield, J. & Milly, P.C. (2017). Divergent surface and total soil moisture projections under global warming. *Geophysical Research Letters*, 44, 236–244.

Bokhorst, S., Huiskes, A., Aerts, R., Convey, P., Cooper, E.J., Dalen, L., Erschbamer, B., Gudmundsson, J., Hofgaard, A., Hollister, R.D. *et al.* (2013). Variable temperature effects of open top chambers at polar and alpine sites explained by irradiance and snow depth. *Global Change Biology*, 19, 64–74.

Canadell, J., Jackson, R., Ehleringer, J., Mooney, H., Sala, O. & Schulze, E.D. (1996). Maximum rooting depth of vegetation types at the global scale. *Oecologia*, 108, 583–595.

Carey, J.C., Tang, J., Templer, P.H., Kroeger, K.D., Crowther, T.W., Burton, A.J., Dukes, J.S., Emmett, B., Frey, S.D., Heskel, M.A. *et al.* (2016). Temperature response of soil respiration largely unaltered with experimental warming. *Proceedings of the National Academy of Sciences*, 113, 13797–13802.

Carlson, R.W. & Bazzaz, F.A. (1982). Photosynthetic and growth response to fumigation with so2 at elevated co2 for c3 and c4 plants. *Oecologia*, 54, 50–54.

Chuine, I., Bonhomme, M., Legave, J.M., García de Cortázar-Atauri, I., Charrier, G., Lacointe, A. & Améglio, T. (2016). Can phenological models predict tree phenology accurately in the future? the unrevealed hurdle of endodormancy break. *Global Change Biology*, 22, 3444–3460.

Chuine, I., Morin, X., Sonié, L., Collin, C., Fabreguettes, J., Degueldre, D., Salager, J.L. & Roy, J. (2012). Climate change might increase the invasion potential of the alien c4 grass setaria parviflora (poaceae) in the mediterranean basin. *Diversity and Distributions*, 18, 661–672.

Clark, J.S., Melillo, J., Mohan, J. & Salk, C. (2014a). The seasonal timing of warming that controls onset of the growing season. *Global Change Biology*, 20, 1136–1145.

Clark, J.S., Salk, C., Melillo, J. & Mohan, J. (2014b). Tree phenology responses to winter chilling, spring warming, at north and south range limits. *Functional Ecology*, 28, 1344–1355.

Cleland, E.E., Chiariello, N.R., Loarie, S.R., Mooney, H.A. & Field, C.B. (2006). Diverse responses of phenology to global changes in a grassland ecosystem. *Proceedings of the National Academy of Sciences of the United States of America*, 103, 13740–13744.

Cox, P.M., Betts, R.A., Jones, C.D., Spall, S.A. & Totterdell, I.J. (2000). Acceleration of global warming due to carbon-cycle feedbacks in a coupled climate model. *Nature*, 408, 184–187.

Dai, A. (2013). Increasing drought under global warming in observations and models. *Nature Climate Change*, 3, 52–58.

Dayton, P.K. (1971). Competition, disturbance, and community organization: the provision and subsequent utilization of space in a rocky intertidal community. *Ecological Monographs*, 41, 351–389.

De Boeck, H.J., Vicca, S., Roy, J., Nijs, I., Milcu, A., Kreyling, J., Jentsch, A., Chabbi, A., Campioli, M., Callaghan, T. *et al.* (2015). Global change experiments: challenges and opportunities. *BioScience*, 65, 922–931.

Diamond, S.E., Penick, C.A., Pelini, S.L., Ellison, A.M., Gotelli, N.J., Sanders, N.J. & Dunn, R.R. (2013). Using physiology to predict the responses of ants to climatic warming. *Integrative and comparative biology*, 53, 965–974.

Dukes, J.S. & Mooney, H.A. (1999). Does global change increase the success of biological invaders? *Trends in Ecology and Evolution*, 14, 135–139.

Dunne, J.A., Harte, J. & Taylor, K.J. (2003). Subalpine meadow flowering phenology responses to climate change: integrating experimental and gradient methods. *Ecological Monographs*, 73, 69–86.

Elmendorf, S.C., Henry, G.H., Hollister, R.D., Björk, R.G., Bjorkman, A.D., Callaghan, T.V., Collier, L.S., Cooper, E.J., Cornelissen, J.H., Day, T.A. *et al.* (2012). Global assessment of experimental climate warming on tundra vegetation: heterogeneity over space and time. *Ecology letters*, 15, 164–175.

Essiamah, S. & Eschrich, W. (1986). Water uptake in deciduous trees during winter and the role of conducting tissues in spring reactivation. *IAWA Journal*, 7, 31–38.

Ettinger, A. & Wolkovich, E. (2018). Microclimate from climate change experiments (MC3E). doi:10.5063/F1QV3JQR.

Famiglietti, J., Devereaux, J., Laymon, C., Tsegaye, T., Houser, P., Jackson, T., Graham, S., Rodell, M. & Oevelen, P.V. (1999). Ground-based investigation of soil moisture variability within remote sensing footprints during the southern great plains 1997 (sgp97) hydrology experiment. *Water Resources Research*, 35, 1839–1851.

Farnsworth, E., Nunez-Farfan, J., Careaga, S. & Bazzaz, F. (1995). Phenology and growth of three temperate forest life forms in response to artificial soil warming. *Journal of Ecology*, 83, 967–977.

Field, C.B., Lobell, D.B., Peters, H.A. & Chiariello, N.R. (2007). Feedbacks of terrestrial ecosystems to climate change. *Annual Review of Environment and Resources*, 32, 1–29.

Franklin, J.F. (1989). Importance and justification of long-term studies in ecology. In: *Long-term studies in ecology*. Springer, pp. 3–19.

Friedlingstein, P., Meinshausen, M., Arora, V.K., Jones, C.D., Anav, A., Liddicoat, S.K. & Knutti, R. (2014). Uncertainties in CMIP5 climate projections due to carbon cycle feedbacks. *Journal of Climate*, 27, 511–526.

Giasson, M.A., Ellison, A.M., Bowden, R., Crill, P.M., Davidson, E., Drake, J., Frey, S., Hadley, J., Lavine, M., Melillo, J. *et al.* (2013). Soil respiration in a northeastern US temperate forest: a 22-year synthesis. *Ecosphere*, 4, 1–28.

Gruner, D.S., Bracken, M.E., Berger, S.A., Eriksson, B.K., Gamfeldt, L., Matthiessen, B., Moorthi, S., Sommer, U. & Hillebrand, H. (2017). Effects of experimental warming on biodiversity depend on ecosystem type and local species composition. *Oikos*, 126, 8–17.

Hagedorn, F., Martin, M., Rixen, C., Rusch, S., Bebi, P., Zürcher, A., Siegwolf, R.T., Wipf, S., Escape, C., Roy, J. *et al.* (2010). Short-term responses of ecosystem carbon fluxes to experimental soil warming at the Swiss alpine treeline. *Biogeochemistry*, 97, 7–19.

Hampe, A. (2004). Bioclimate envelope models: what they detect and what they hide. *Global Ecology and Biogeography*, 13, 469–471.

Harte, J., Saleska, S.R. & Levy, C. (2015). Convergent ecosystem responses to 23-year ambient and manipulated warming link advancing snowmelt and shrub encroachment to transient and long-term climate–soil carbon feedback. *Global Change Biology*, 21, 2349–2356.

Harte, J., Torn, M.S., Chang, F.R., Feifarek, B., Kinzig, A.P., Shaw, R. & Shen, K. (1995). Global warming and soil microclimate: Results from a meadow-warming experiment. *Ecological Applications*, 5, 132–150.

Helm, B. & Shavit, A. (2017). *Dissecting and reconstructing time and space for replicable biological research*, New Haven, CT: Yale University Press, pp. 233–249.

Hobbie, S.E., Shevtsova, A. & Chapin III, F.S. (1999). Plant responses to species removal and experimental warming in Alaskan tussock tundra. *Oikos*, 84, 417–434.

Hoeppner, S.S. & Dukes, J.S. (2012). Interactive responses of old-field plant growth and composition to warming and precipitation. *Global Change Biology*, 18, 1754–1768.

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

Hurlbert, S.H. (1984). Pseudoreplication and the design of ecological field experiments. *Ecological monographs*, 54, 187–211.

Huston, M.A. (1997). Hidden treatments in ecological experiments: re-evaluating the ecosystem function of biodiversity. *Oecologia*, 110, 449–460.

Ibanez, I., Clark, J.S., Dietze, M.C., Feeley, K., Hersh, M., LaDeau, S., McBride, A., Welch, N.E. & Wolosin, M.S. (2006). Predicting biodiversity change: Outside the climate envelope, beyond the species-area curve. *Ecology*, 87, 1896–1906.

Johnson, P.D. & Besselsen, D.G. (2002). Practical aspects of experimental design in animal research. *ILAR Journal*, 43, 202–206.

Kearney, M. & Porter, W. (2009). Mechanistic niche modelling: combining physiological and spatial data to predict species' ranges. *Ecology Letters*, 12, 334–350.

Kennedy, A. (1995). Temperature effects of passive greenhouse apparatus in high-latitude climate change experiments. *Functional Ecology*, 9, 340–350.

Kharouba, H.M., Vellend, M., Sarfraz, R.M. & Myers, J.H. (2015). The effects of experimental warming on the timing of a plant–insect herbivore interaction. *Journal of Animal Ecology*, 84, 785–796.

Kimball, B. (2005). Theory and performance of an infrared heater for ecosystem warming. *Global Change Biology*, 11, 2041–2056.

Kimball, B.A., Conley, M.M., Wang, S., Lin, X., Luo, C., Morgan, J. & Smith, D. (2008). Infrared heater arrays for warming ecosystem field plots. *Global Change Biology*, 14, 309–320.

Marchin, R.M., Salk, C.F., Hoffmann, W.A. & Dunn, R.R. (2015). Temperature alone does not explain phenological variation of diverse temperate plants under experimental warming. *Global Change Biology*, 21, 3138–3151.

Marion, G., Henry, G., Freckman, D., Johnstone, J., Jones, G., Jones, M., Levesque, E., Molau, U., Mølgaard, P., Parsons, A. *et al.* (1997). Open-top designs for manipulating field temperature in high-latitude ecosystems. *Global Change Biology*, 3, 20–32.

McDaniel, M., Kaye, J. & Kaye, M. (2014a). Do "hot moments" become hotter under climate change? soil nitrogen dynamics from a climate manipulation experiment in a post-harvest forest. *Biogeochemistry*, 121, 339–354.

McDaniel, M., Wagner, R., Rollinson, C., Kimball, B., Kaye, M. & Kaye, J. (2014b). Microclimate and ecological threshold responses in a warming and wetting experiment following whole tree harvest. *Theoretical and Applied Climatology*, 116, 287–299.

Melillo, J., Frey, S., DeAngelis, K., Werner, W., Bernard, M., Bowles, F., Pold, G., Knorr, M. & Grandy, A. (2017). Long-term pattern and magnitude of soil carbon feedback to the climate system in a warming world. *Science*, 358, 101–105.

Menke, S.B., Harte, J. & Dunn, R.R. (2014). Changes in ant community composition caused by 20 years of experimental warming vs. 13 years of natural climate shift. *Ecosphere*, 5, 1–17.

Milcu, A., Puga-Freitas, R., Ellison, A.M., Blouin, M., Scheu, S., Girin, T., Frechet, G., Rose, L., Scherer-Lorenzen, M., Barot, S. *et al.* (2016). Systematic variability enhances the reproducibility of an ecological study. *bioRxiv(beta)*, p. 080119.

Moise, E.R. & Henry, H.A. (2010). Like moths to a street lamp: exaggerated animal densities in plot-level global change field experiments. *Oikos*, 119, 791–795.

Moorcroft, P.R. (2006). How close are we to a predictive science of the biosphere? *Trends in Ecology & Evolution*, 21, 400–407.

Morin, X., Roy, J., Sonié, L. & Chuine, I. (2010). Changes in leaf phenology of three European oak species in response to experimental climate change. *New Phytologist*, 186, 900–910.

Ohlemüller, R., Gritti, E.S., Sykes, M.T. & Thomas, C.D. (2006). Towards European climate risk surfaces: the extent and distribution of analogous and non-analogous climates 1931–2100. *Global Ecology and Biogeography*, 15, 395–405.

Parmesan, C. (2006). Ecological and evolutionary responses to recent climate change. *Annual Review of Ecology Evolution and Systematics*, 37, 637–669.

Pearson, R.G. & Dawson, T.P. (2004). Bioclimate envelope models: what they detect and what they hide - response to hampe (2004). *Global Ecology and Biogeography*, 13, 471–473.

Pelini, S., Diamond, S., Nichols, L., Stuble, K., Ellison, A.M., Sanders, N., Dunn, R. & Gotelli, N. (2014). Geographic differences in effects of experimental warming on ant species diversity and community composition. *Ecosphere*, 5, 1–12.

Pelini, S.L., Bowles, F.P., Ellison, A.M., Gotelli, N.J., Sanders, N.J. & Dunn, R.R. (2011). Heating up the forest: open-top chamber warming manipulation of arthropod communities at Harvard and Duke Forests. *Methods in Ecology and Evolution*, 2, 534–540.

Peterjohn, W.T., Melillo, J.M., Bowles, F.P. & Steudler, P.A. (1993). Soil warming and trace gas fluxes: experimental design and preliminary flux results. *Oecologia*, 93, 18–24.

Polade, S.D., Pierce, D.W., Cayan, D.R., Gershunov, A. & Dettinger, M.D. (2014). The key role of dry days in changing regional climate and precipitation regimes. *Scientific reports*, 4, 4364.

Polgar, C.A., Primack, R.B., Dukes, J.S., Schaaf, C., Wang, Z. & Hoeppner, S.S. (2014). Tree leaf out response to temperature: comparing field observations, remote sensing, and a warming experiment. *International journal of biometeorology*, 58, 1251–1257.

Price, M.V. & Waser, N.M. (1998). Effects of experimental warming on plant reproductive phenology in a subalpine meadow. *Ecology*, 79, 1261–1271.

Quinn, G.P. & Keough, M.J. (2002). *Experimental design and data analysis for biologists*. Cambridge University Press.

Reich, P.B., Hobbie, S.E., Lee, T.D. & Pastore, M.A. (2018). Unexpected reversal of C3 versus C4 grass response to elevated CO2 during a 20-year field experiment. *Science*, 360, 317–320.

Reich, P.B., Sendall, K.M., Rice, K., Rich, R.L., Stefanski, A., Hobbie, S.E. & Montgomery, R.A. (2015). Geographic range predicts photosynthetic and growth response to warming in co-occurring tree species. *Nature Climate Change*, 5, 148–152.

Reich, P.B., Sendall, K.M., Stefanski, A., Wei, X., Rich, R.L. & Montgomery, R.A. (2016). Boreal and temperate trees show strong acclimation of respiration to warming. *Nature*, 531, 633–636.

Rich, R.L., Stefanski, A., Montgomery, R.A., Hobbie, S.E., Kimball, B.A. & Reich, P.B. (2015). Design and performance of combined infrared canopy and belowground warming in the B4WarmED (boreal forest warming at an ecotone in danger) experiment. *Global change biology*, 21, 2334–2348.

Rollinson, C.R. & Kaye, M.W. (2012). Experimental warming alters spring phenology of certain plant functional groups in an early successional forest community. *Global Change Biology*, 18, 1108–1116.

Rollinson, C.R. & Kaye, M.W. (2015). Modeling monthly temperature in mountainous ecoregions: importance of spatial scale for ecological research. *Climate Research*, 64, 99–110.

Saleska, S.R., Shaw, M.R., Fischer, M.L., Dunne, J.A., Still, C.J., Holman, M.L. & Harte, J. (2002). Plant community composition mediates both large transient decline and predicted long-term recovery of soil carbon under climate warming. *Global Biogeochemical Cycles*, 16, 3–1–3–18.

Seager, R., Neelin, D., Simpson, I., Liu, H., Henderson, N., Shaw, T., Kushnir, Y., Ting, M. & Cook, B. (2014). Dynamical and thermodynamical causes of large-scale changes in the hydrological cycle over north america in response to global warming. *Journal of Climate*, 27, 7921–7948.

Seager, R., Ting, M., Li, C., Naik, N., Cook, B., Nakamura, J. & Liu, H. (2013). Projections of declining surface-water availability for the southwestern United States. *Nature Climate Change*, 3, 482.

Shaver, G.R., Canadell, J., Chapin, F.S., Gurevitch, J., Harte, J., Henry, G., Ineson, P., Jonasson, S., Melillo, J., Pitelka, L. *et al.* (2000). Global warming and terrestrial ecosystems: A conceptual framework for analysis ecosystem responses to global warming will be complex and varied. Ecosystem warming experiments hold great potential for providing insights on ways terrestrial ecosystems will respond to upcoming decades of climate change. Documentation of initial conditions provides the context for understanding and predicting ecosystem responses. *BioScience*, 50, 871–882.

Sheldon, K.S., Yang, S. & Tewksbury, J.J. (2011). Climate change and community disassembly: impacts of warming on tropical and temperate montane community structure. *Ecology Letters*, 14, 1191–1200.

Sherry, R.A., Zhou, X., Gu, S., 3rd, J.A.A., Schimel, D.S., Verburg, P.S., Wallace, L.L. & Luo, Y. (2007). Divergence of reproductive phenology under climate warming. *Proceedings of the National Academy of Sciences of the United States of America*, 104, 198–202.

Shukla, J. & Mintz, Y. (1982). Influence of land-surface evapotranspiration on the earth's climate. *Science*, 215, 1498–1501.

Shuman, B.N. & Burrell, S.A. (2017). Centennial to millennial hydroclimatic fluctuations in the humid northeast United States during the holocene. *Quaternary Research*, 88, 1–11.

Spector, R. (2001). Progress in the search for ideal drugs. *Pharmacology*, 64, 1–7.

Stocker, T., Qin, D., Plattner, G., Tignor, M., Allen, S., Boschung, J., Nauels, A., Xia, Y., Bex, B. & Midgley, B. (2013). IPCC 2013. Climate change 2013: The physical science basis. Contribution of Working Group II to the fifth assessment report of the Intergovernmental Panel on Climate Change.

Swab, R.M., Regan, H.M., Keith, D.A., Regan, T.J. & Ooi, M.K.J. (2012). Niche models tell half the story: spatial context and life-history traits influence species responses to global change. *Journal of Biogeography*, 39, 1266–1277.

Tamaki, G., Weiss, M.A. & Long, G.E. (1981). Evaluation of plant density and temperature in predator-prey interactions in field cages. *Environmental Entomology*, 10, 716–720.

Templer, P.H., Phillips, N.G., Ellison, A.M. & Pelini, S.L. (2016). Ecosystem warming increases sap flow rates of northern red oak trees. *Ecosphere*, 7.

Teuling, A.J. & Troch, P.A. (2005). Improved understanding of soil moisture variability dynamics. *Geophysical Research Letters*, 32.

Thomas, C.D., Cameron, A., Green, R.E., Bakkenes, M., Beaumont, L.J., Collingham, Y.C., Erasmus, B.F.N., de Siqueira, M.F., Grainger, A., Hannah, L., Hughes, L., Huntley, B., van Jaarsveld, A.S., Midgley, G.F., Miles, L., Ortega-Huerta, M.A., Peterson, A.T., Phillips, O.L. & Williams, S.E. (2004). Extinction risk from climate change. *Nature*, 427, 145–148.

Thorne, P., Donat, M., Dunn, R., Williams, C., Alexander, L., Caesar, J., Durre, I., Harris, I., Hausfather, Z., Jones, P. *et al.* (2016). Reassessing changes in diurnal temperature range: Intercomparison and evaluation of existing global data set estimates. *Journal of Geophysical Research: Atmospheres*, 121, 5138–5158.

Thuiller, W. (2004). Patterns and uncertainties of species' range shifts under climate change. *Global Change Biology*, 10, 2020–2027.

Urban, M.C., Tewksbury, J.J. & Sheldon, K.S. (2012). On a collision course: competition and dispersal differences create no-analogue communities and cause extinctions during climate change. *Proceedings of the Royal Society B-Biological Sciences*, 279, 2072–2080.

Vasseur, D.A., DeLong, J.P., Gilbert, B., Greig, H.S., Harley, C.D., McCann, K.S., Savage, V., Tunney, T.D. & O'Connor, M.I. (2014). Increased temperature variation poses a greater risk to species than climate warming. *Proceedings of the Royal Society of London B: Biological Sciences*, 281, 20132612.

Vose, R.S., Easterling, D.R. & Gleason, B. (2005). Maximum and minimum temperature trends for the globe: An update through 2004. *Geophysical Research Letters*, 32, e01221.

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

Williams, J.W., Jackson, S.T. & Kutzbacht, J.E. (2007). Projected distributions of novel and disappearing climates by 2100 AD. *Proceedings of the National Academy of Sciences of the United States of America*, 104, 5738–5742.

Wolkovich, E.M., Cook, B.I., Allen, J.M., Crimmins, T.M., Betancourt, J.L., Travers, S.E., Pau, S. *et al.* (2012). Warming experiments underpredict plant phenological responses to climate change. *Nature*, 485, 494–497.

Woodward, F. (1992). Predicting plant responses to global environmental change. *New Phytologist*, 122, 239–251.

Zhu, K., Chiariello, N.R., Tobeck, T., Fukami, T. & Field, C.B. (2016). Nonlinear, interacting responses to climate limit grassland production under global change. *Proceedings of the National Academy of Sciences*, 113, 10589–10594.

Figures

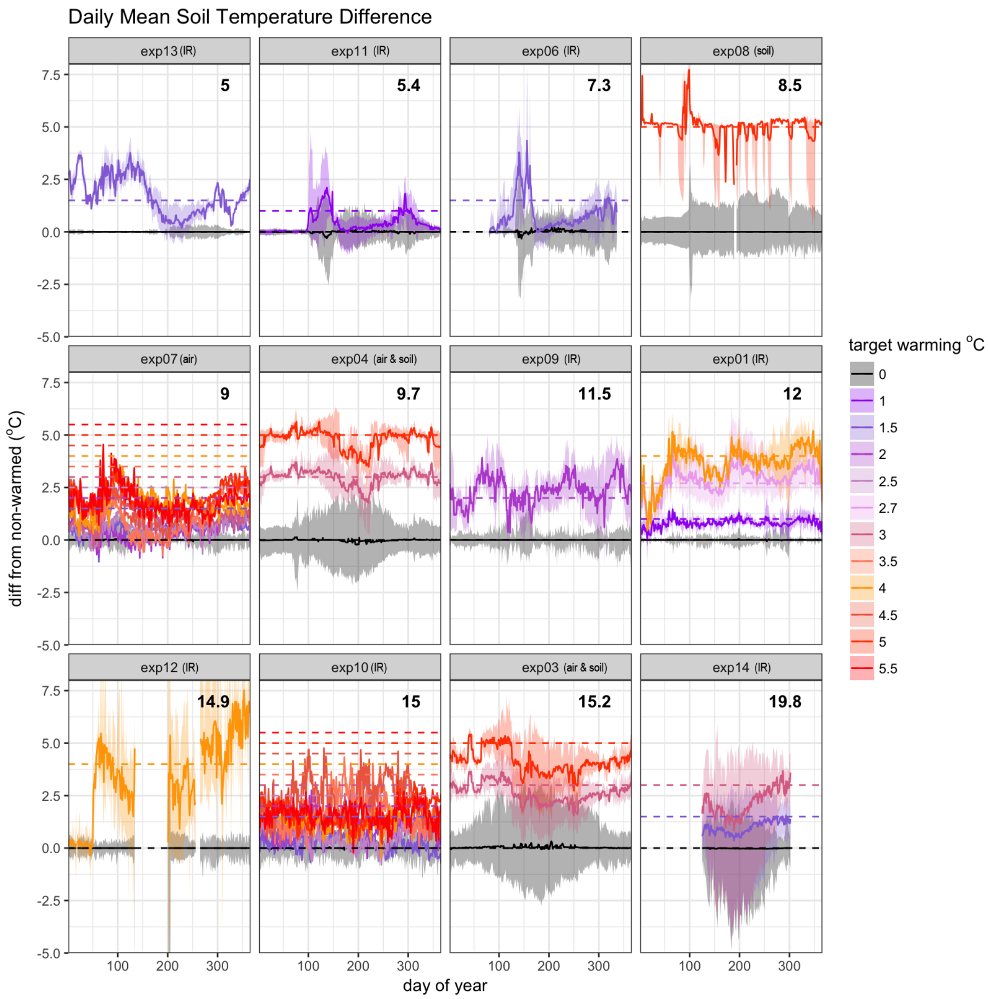

Figure 1: **Deviations in daily observed warming from mean control soil temperature for 12 study sites,** excluding data from plots that manipulated precipitation. We show soil, rather than above-ground, temperature, as this was the most frequently recorded temperature variable in the MC3E database. Solid lines show observed difference between warming treatment (colors) and control (black) plots, averaged across replicates and years; shading shows 95% confidence intervals. Dashed lines represent target warming levels. (Note that the following sites had no explicit target temperature: exp06, exp11, exp12; in exp01, only the highest warming treatment had a target temperature; for these studies and treatments, we used their reported level of warming.) Two sites not shown here did not monitor soil temperature. Sites are ordered by low to high mean annual soil temperature (shown in the upper right corner of each panel). The heating type is listed in parentheses next to the site number (IR= infrared, soil= soil cables, air= forced air).

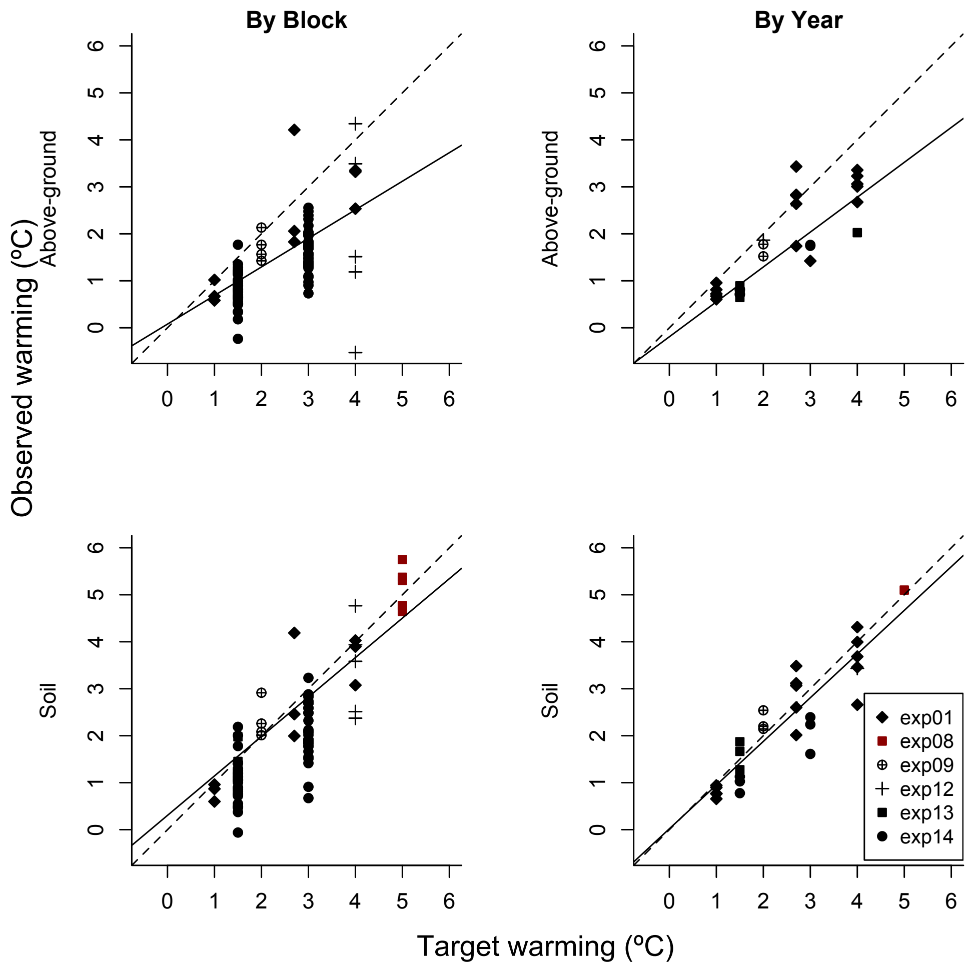

Figure 2: **Observed warming over space and time, for above-ground and soil temperatures,** excluding data from plots that manipulated precipitation. Above-ground temperature includes air, canopy, and surface temperature. Points represent the difference between treatment and control plots by block (i.e., one data point per block) and by year (i.e., one data point per year). The solid line is the fitted relationship between observed and target warming and the dashed line shows when observed warming is exactly equal to target warming (1:1). Black symbols represent studies using infrared; red represents soil warming cables (only exp08); no studies with forced air heating used a blocked design. Note that the following studies had no explicit target temperature: exp06, exp11, exp12; for these studies, we used their reported level of warming. For exp01, only the treatment with the greatest warming had a target temperature. See Supplemental Materials (especially Tables S5 and S6) for details.

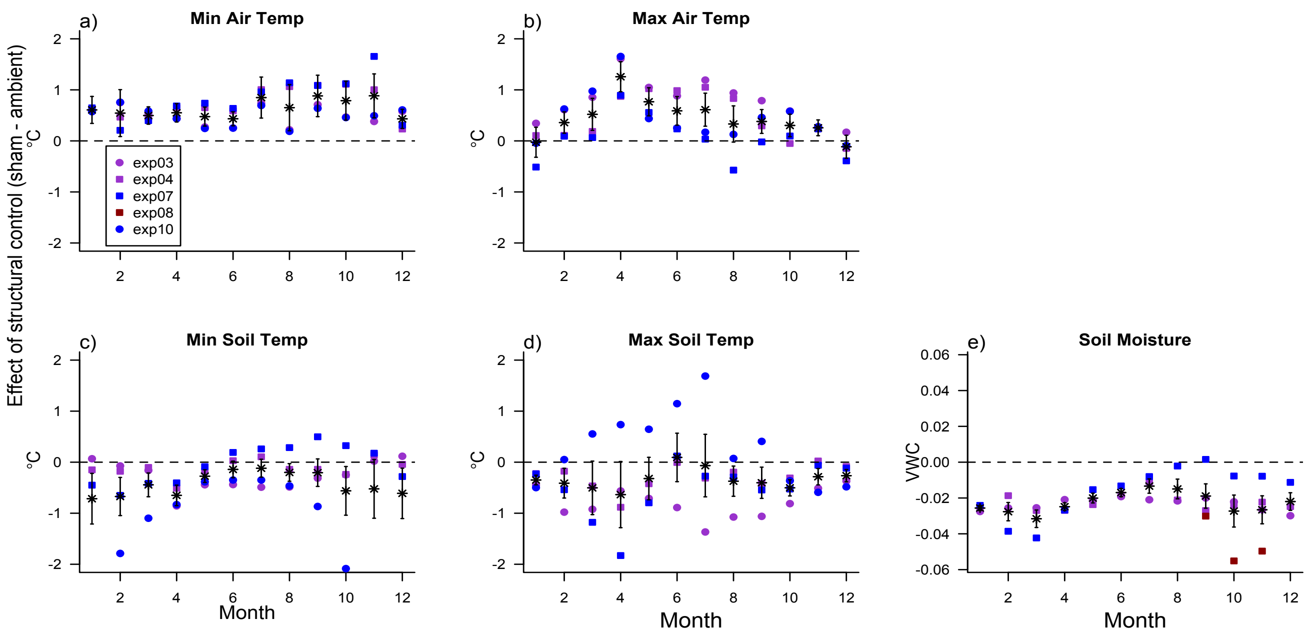

Figure 3: **Deviations in measured abiotic variables by month in structural controls compared to ambient controls** (i.e., with no control chambers or warming infrastructure in place). Above-ground temperatures (which include includes air, canopy, and surface temperatures) were higher (a,b), whereas soil temperature (c,d) and soil moisture (e) were lower in structural controls compared with ambient controls. We show overall (fixed) effects in black from monthly mixed-effects models; site-level random effects are shown by squares (for the three studies conducted at Harvard Forest in Massachusetts, USA) and circles (the two studies conducted at Duke Forest in North Carolina, USA). Colors vary by heating type: red represents soil warming cables, blue represents forced air; purple represents combined soil warming cables and forced air heating (no studies with infrared heating included both control types). See Supplemental Materials for details (Tables S8-S13).

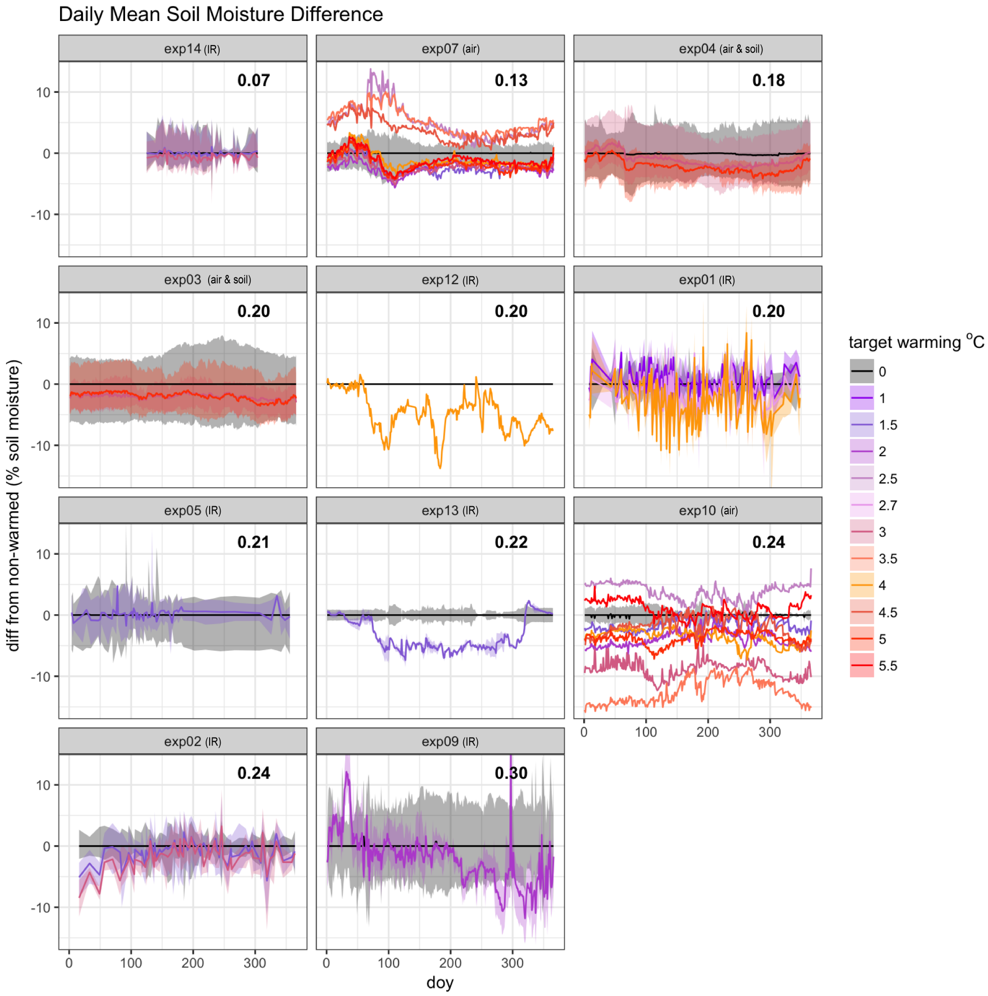

Figure 4: **Deviations in daily observed soil moisture,** shown for the 11 study sites that continuously monitored soil moisture, excluding data from plots that manipulated precipitation. Black lines represent control plots, and colored lines represent warming treatments with various target warming levels (or reported warming, if there was no explicit target temperature). The number of temperature treatment levels vary from one (e.g., exp08, exp11) to nine (exp07 and exp10, which used an unreplicated regression design). Sites are ordered by low to high mean annual soil moisture (shown in the upper right corner of each plot). All experiments measured soil moisture in volumetric water content, as a percentage of the soil volume in the sample, scaled from 0 to 100; the absolute difference between treatment and control plots is shown.

.

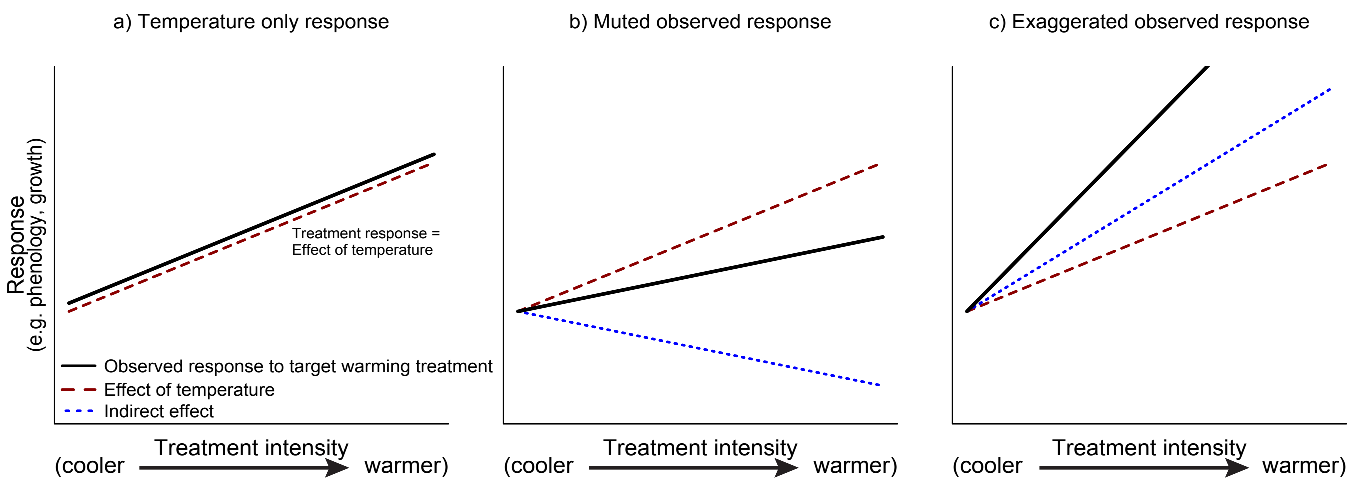

Figure 5: **Theoretical biological responses to experimental warming and their interpretation**. Direct responses to temperature alone (a) can be easily understood. Complications arise when biological responses are a mix of the direct temperature and indirect non-temperature effects of experimental warming. Then experimental warming may cause biological responses to be muted (b) or exaggerated (c). Quantifying, interpreting, and reporting these non-temperature effects in experiments is critical, and their presence is both a strength and a challenge of climate change experiments. They may represent ecologically realistic effects that might not have been predicted without the *in situ* field experiment. Alternatively, they may represent artifacts that are unlikely to occur outside of an experimental context. Slopes of these example lines assume a linear response with additive direct and indirect effects. The relationship between these effects could be more complex (e.g., nonlinear; antagonistic, multiplicative, or otherwise interactive).

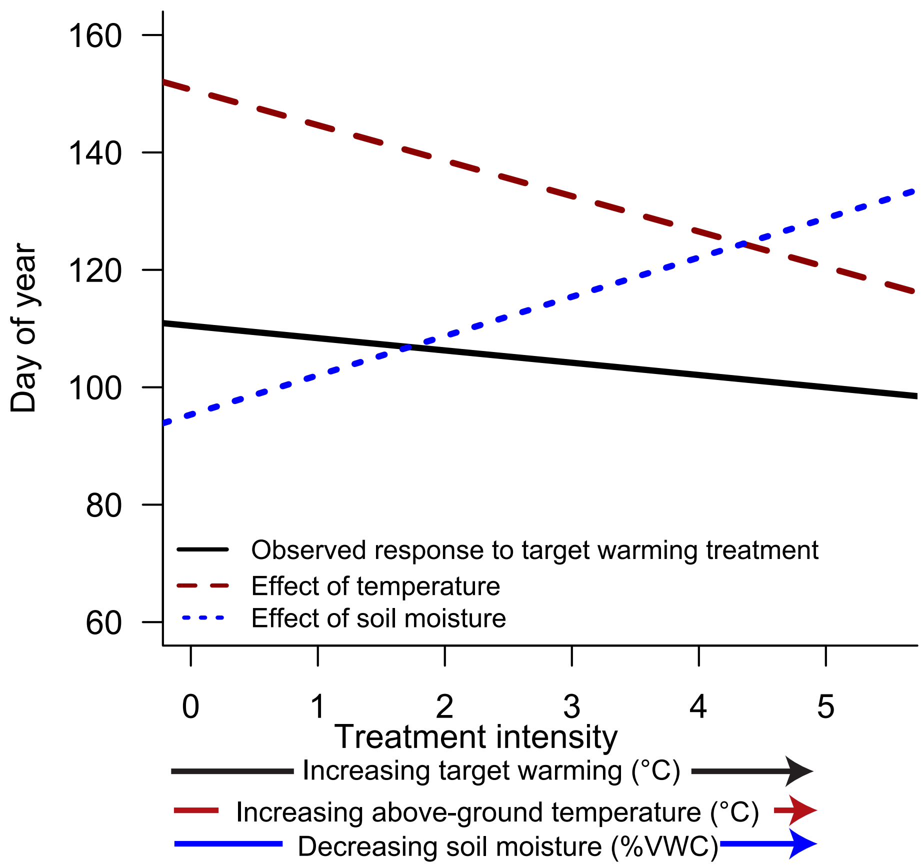

Figure 6: **Observed response of budburst day of year to experimental climate change** is an example of a muted response: the observed response to increasing treatment intensity (i.e., the coefficient of a model fit with only target [or reported, if there was no explicit target] temperature as the explanatory variable, black line; units for x-axis are °C of target warming) suggests a weaker temperature sensitivity than the effect of temperature in a more biologically accurate (and better-fitting) model that includes both measured above-ground temperature (dashed red line, for which x-axis units are °C of measured temperature) and soil moisture (dotted blue line, for which x-axis units are % VWC, decreasing from left to right in conjunction with warming intensity), as well as their interaction. Analysis includes all studies that monitored budburst and measured soil moisture and above-ground temperature (exp01, exp03, exp04, exp07, exp10); structural control data were used for this analysis (ambient controls were excluded from those studies that contained both). See Supplemental Materials, especially Tables S17 & S18, for additional details.