ECOLOGY LETTERS

Ecology Letters, (2019) doi: 10.1111/ele.13223

REVIEW AND SYNTHESIS

How do climate change experiments alter plot-scale climate?

A. K. Ettinger,[1,2*] I. Chuine,[3] B. I. Cook,[4,5] J. S. Dukes,[6] A. M. Ellison,[7] M. R. Johnston,[8] A. M. Panetta,[9] C. R. Rollinson,[10] Y. Vitasse[11,12] and E. M. Wolkovich[1,8,13]

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 to compile a database of daily plot-scale climate data from 15 active-warming experiments. We find that the common practices of analysing treatments as mean or categorical changes (e.g. warmed vs. 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 within a study, differed by up to 1.6 °C (63% of target), on average, across six studies with blocked designs. Variation was high across sites and designs: for example, plots differed by 1.1 °C (47% of target) on average, for infrared studies with feedback control (n = 3) vs. by 2.2 °C (80% of target) on average for infrared with constant wattage designs (n = 2). Warming treatments produce non-temperature effects as well, such as soil drying. The combination of these direct and indirect effects is complex and can have important biological consequences. With a case study of plant phenology across five experiments in our database, we show how accounting for drier soils with warming tripled the estimated sensitivity of budburst to temperature. We provide recommendations for future analyses, experimental design, and data sharing to improve our mechanistic understanding from climate change experiments, and thus their utility to accurately forecast species' responses.

Keywords

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

Ecology Letters (2019)

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 climate, and process-based modelling; 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. Hampe 2004; Pearson & Dawson 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 &

[1]Arnold Arboretum of Harvard University, Boston, MA 02131, USA
[2]Tufts University, Medford, MA 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, NY 10964, USA
[5]NASA Goddard Institute for Space Studies, New York, NY 10025, USA
[6]Department of Forestry and Natural Resources and Department of Biological Sciences, Purdue University, West Lafayette, IN 47907, USA
[7]Harvard Forest, Harvard University, Petersham, MA 01366, USA
[8]Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA 02138, USA
[9]Department of Ecology and Evolutionary Biology, University of Colorado, Boulder, CO 80309, USA
[10]Center for Tree Science, The Morton Arboretum, Lisle, IL 60532, USA
[11]Swiss Federal Institute for Forest, Snow and Landscape Research WSL, Birmensdorf, Switzerland
[12]SwissForestLab, Birmensdorf, Switzerland
[13]Forest & Conservation Sciences, Faculty of Forestry, University of British Columbia, Vancouver, BC, V6T 1Z4, Canada
*Correspondence: E-mail: aettinger@fas.harvard.edu

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 parameterise 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) 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 active-warming treatments alter microclimate, both directly and indirectly, across multiple studies. The qualitative challenges and opportunities provided by climate change experiments have been summarised 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; Pelini *et al.* 2011; McDaniel *et al.* 2014b). However, our quantitative meta-analysis allows us to examine trends across sites and warming designs (Box 1, Table 1), 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, may lead to inaccurate quantification of plant sensitivity to temperature shifts. Finally, we synthesise our findings to make recommendations for future analysis and design of climate change experiments (Box 2).

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 analysed extensively already and are known to have distinct issues, including reduction in wind, overheating, and high 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 data sets 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 (Fig. 1; Fig. S1, Table S1), which is available at KNB (Ettinger & Wolkovich 2018). We examined how these experiments altered microclimate, using mixed-effects models to estimate across-study effects, while also accounting for inherent differences among studies (through a random effect of study on the intercept).

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.

Box 1 Different methods for achieving warming

Active-warming experiments may differ both in the way that they achieve warming ('warming type' in Table S1), and the way that warming is controlled ('warming control' in Table S1). There are three warming types used by studies in the MC3E database (Fig. 2). These are infrared (n = 9), an open-air (chamber-less) method in which infrared heaters are mounted above the ground; forced air (n = 3), in which air is heated and then pumped through an airflow system into a chamber; and soil warming (n = 1), a chamber-less method in which soil is heated with buried electric resistance cable. Two additional studies in the database used combined forced air in chambers and soil warming. Warming is controlled by either constant wattage, in which an unvarying energy output is used, or by feedback control, in which energy outputs are linked to a thermometer and varied depending on the measured temperature in plots, in order to maintain consistent warming levels.

In this paper, we describe complications and non-temperature effects associated with active-warming experiments, across these divergent warming methodologies. Some of the non-temperature effects described may be more likely to occur with particular methods. Alterations to airflow, for example, may be most dramatic with methods employing chambers. Plot shading and precipitation interference are likely to occur in chamber and infrared techniques, which both involve above-ground infrastructure, and less likely in methods that only warm from the soil. Soil warming methods, however, may be less representative of climate change, which will be driven by above-ground rather than below-ground warming. Warming cables also disturb the soil, potentially altering conductivity, water flow, and other soil properties. The biological impacts of such effects may be further enhanced or muted based on site characteristics (e.g. if a site is already heavily shaded, impacts from above-ground infrastructure shading may be lower).

Table 1 highlights that there may be differences in non-temperature effects across these different warming methodologies. In the MC3E database, sample sizes within each warming and control type were quite low, so we were unable to statistically distinguish differences in non-temperature effects across the different methods in all analyses. For example, the constant wattage control studies have greater average variation (2.2 °C variation on average for constant vs. 1.1 °C variation for feedback), but this difference is not significant (P = 0.21). We note that the studies showing both the greatest and least variation employed constant wattage (greatest: plots in exp12, with target warming of 4.0 °C, had mean warming levels that varied by as much as 4.9 °C; least: plots in exp13 with 1.5 °C of target warming, varied by 0.03 °C). These results are not conclusive, because our sample size is quite low (n = 3 studies for constant and n = 2 studies for feedback studies with blocked designs).

We expect that the list of non-temperature effects in Table 1 is not exhaustive, but represents what we can document here or has been documented previously. We recommend additional detailed studies of these, and other, effects across warming designs. This will allow future researchers to more fully evaluate the challenges and opportunities of each method, and select an experimental approach well-suited to their particular research focus.

weekly, seasonally or annually). Not surprisingly then, when analysing 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; Pelini *et al.* 2011; McDaniel *et al.* 2014b). 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. When possible, we compare and contrast these factors across different study methodologies, such as infrared warming vs. forced air chambers and constant wattage vs. feedback control, because effects on microclimate may vary across these different methodologies (Fig. 2, Box 1).

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 Fig. 3 to Fig. 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 vs. forced air) or warming control (feedback vs. constant). Soil daily temperature range was not affected by experimental warming, as warming altered minimum and maximum daily temperatures similarly (Table S4).

Table 1 Summary of measured warming and documented non-temperature effects, by warming technique, for studies included in the MC3E database

Warming type	Warming control	Study	Target (min–max)	Above-ground mean (SE)	Above-ground range	Soil mean (SE)	Soil range	*n*	Soil drying	Other non-temperature effects
Infrared	Constant	exp05,06,11–14	2.2 (1.2–4)	0.52 (0.03)	−5.22–2.16	0.68 (0.02)	−0.07–1.56	6	Fig. 4, Tables S15–S16, Kimball (2005)	+ shading (Kimball *et al.* 2008)
Infrared	Feedback	exp01,02,09	2.5 (1–4)	0.84 (0.05)	−0.19–1.86	0.96 (0.05)	−0.03–1.85	3	Fig.4, Tables S15–S16, Kimball (2005), Sherry *et al.* 2007	+ freeze-thaw cycles (McDaniel *et al.* 2014b) + shading (Kimball *et al.* 2008) + VPD (Morin *et al.* 2010)
Forced air	Feedback	exp07,10,15	3.5 (1.5–5.5)	0.95 (0.02)	0.43–1.73	0.36 (0.02)	−0.42–1.06	3	Figs 3–4 Norby *et al.* (1997)	+VPD (Norby *et al.* 1997) Air flow (Norby *et al.* 1997)
Soil cables	Feedback	exp08	5			1.01 (0.02)	0.9–1.08	1	Figs 3–4, Peterjohn *et al.* (1993)	+ CO_2 flux (Peterjohn *et al.* 1993) + N mineralisation (Peterjohn *et al.* 1993)
Force air, soil cables	Feedback	exp03,04	4 (3–5)	0.49 (0.06)	−0.02–0.94	0.75 (0.07)	0.01–1.05	2	Figs 3–4	Air flow (Clark *et al.* 2014b)

Notes. Summaries of the target warming treatments (°C) and measured warming for above-ground temperature and soil temperature are given. Measured warming is standardised per degree of target warming, and is shown here for warming treatments only (precipitation treatments are excluded). Thus, measured warming is the difference between mean annual temperature (MAT) of control plots and MAT of each treatment level within year (and block, if applicable) of a study. Mean difference (with standard error) and the range of differences in warming are shown, across all years (and blocks, if applicable). *n* is the number of studies of each warming technique and control type combination in the MC3E database. 'Soil drying' indicates figures showing an effect of warming on soil water content and references of previous studies reporting such effect.'Other non-temperature effects' indicates other effects that have been studied in individual previous studies for each type of warming, including altered vapor pressure deficit (VPD), nitrogen (N) mineralisation, and other effects. See Table S1 for additional details of studies included in the MC3E database.

Box 2 Recommendations for future climate change experiments

1. *Collect and analyse plot-level climate data*. This includes analysing 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. *Analyse measured climate variables rather than targets*. There can be substantial variation in the effects of warming and precipitation treatments among plots and across time (Fig. 3). Analysing 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 wattage) to aid future analyses of the performance of different designs, across sites and over time (Box 1, Table 1).
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 (Fig. 4).
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. *Maximise 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 (Fig. 3). In addition, long-term responses of individuals and populations can differ from transient responses (Franklin 1989; Saleska *et al*. 2002; 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 versus feedback wattage designs (Box 1).

We observed strong seasonal and annual variations in the effects of experimental warming (Figs 1 and 3, 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 (Fig. 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). Differences between target and actual warming are likely to be particularly great for studies employing constant wattage, rather than feedback control (Box 1, Fig. 2). 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. Hagedorn *et al*. 2010; Clark *et al*. 2014a,b).

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 two infrared with feedback control, three infrared with constant wattage, and one soil warming cable with feedback control experiments. We found that the amount of observed warming frequently varied by more than 1 °C (mean = 1.6 °C, maximum = 4.9°C) among blocks (Fig. 3, Table S6). This variation in warming is substantial, as it is equivalent to the target warming treatment for many studies. It also appears to vary substantially among sites, which differ in warming methodologies and environmental characteristics, though low sample sizes make disentangling the effect of warming method difficult (Box 1). The differences in warming among blocks may be caused by fine-scale variation in vegetation, slope, aspect, soil type, or other

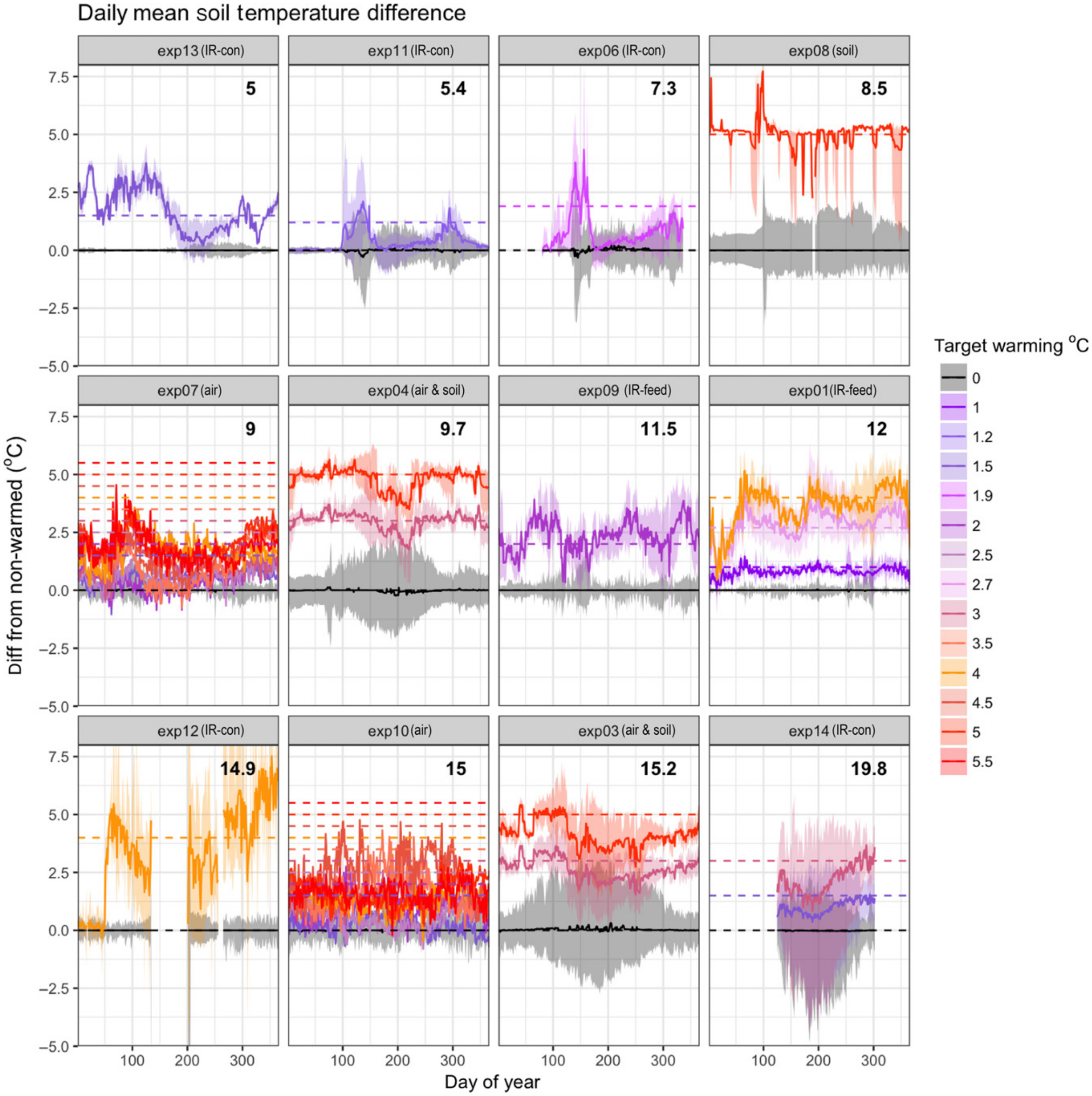

Figure 1 Deviations in daily observed warming from mean control soil temperature for 12 study sites, excluding data from plots that manipulated precipitation. We show soil, rather than above-ground, temperature, as this was the most frequently recorded temperature variable in the MC3E database. Solid lines show observed difference between warming treatment (colours) 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.) Three 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). Heating type is listed in parentheses next to the site number: IR-con = infrared with constant wattage, IR-feed = infrared with feedback control, soil = soil cables, air = forced air.

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 the 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*.

Soil cable warming experiment at Harvard Forest, Petersham, Massachusetts, USA (Melillo et al. 2017). Photo credit: A. Barker-Plotkin.

Forced air chamber warming experiment at Duke Forest, Hillsborough, North Carolina, USA (exp10, Pelini *et al.*, 2011). Photo credit: A. Ellison.

Infrared warming experiment in Pennsylvania, USA (exp09, Rollinson *et al.*, 2012). Photo credit: C. Rollinson.

Infrared warming experiment in Montpelier, France (exp02, Morin *et al.*, 2010). Photo credit: I. Chuine.

Figure 2 Photographs of different warming methodologies used by studies in the MC3E database. See Box 1.

2016). However, because fine-scale and temporal variations in warming treatments are rarely analysed 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 analysed.

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; Pelini *et al.* 2011; Clark *et al.* 2014b; Marchin *et al.* 2015) (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 the warming infrastructure: 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 = 4) or ambient controls (n = 5). 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), and infrastructure effects have been documented in other studies (e.g. shading, Table 1.

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 (Fig. 4a–d). This general pattern was consistent across different temperature models (mean, minimum, and maximum temperatures), although the magnitude varied among seasons, studies, and years (Fig. 4a–d, Tables S8–S11). We also found that experimental infrastructure decreased soil moisture relative to ambient conditions across all seasons, studies, and years (Fig. 4e, Tables S12 and 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, particularly in chamber warming (i.e. 4 of the 5 studies included in the above analysis, see also Aronson & McNulty 2009). Specifically, soil temperatures may be cooler in structural controls for forced air studies because the

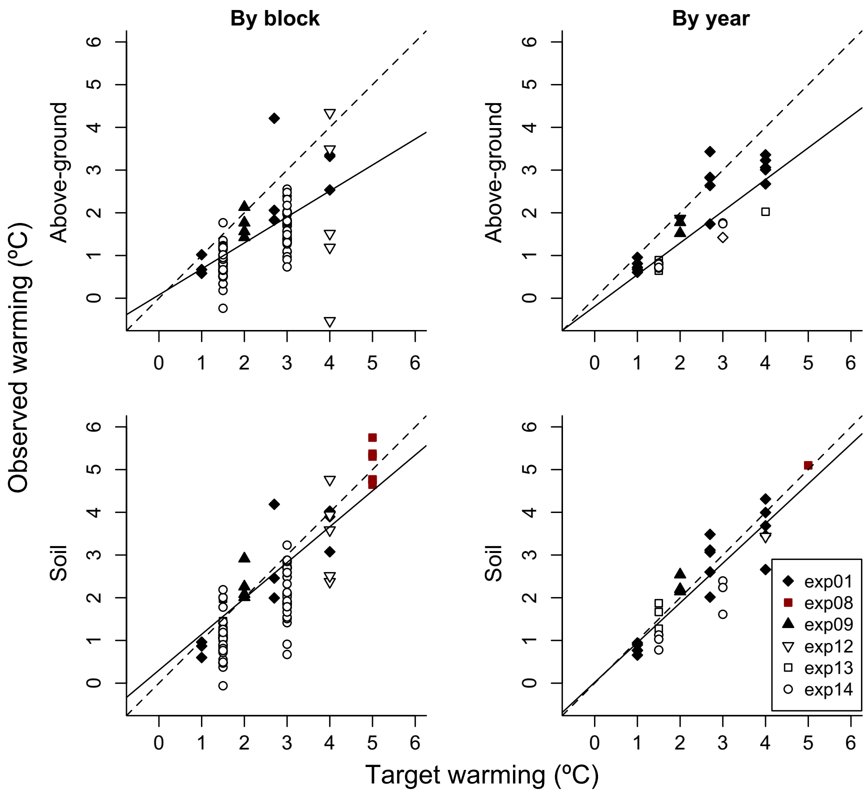

Figure 3 Observed warming over space and time, for above-ground and soil temperatures, excluding data from plots that manipulated precipitation. Above-ground temperature includes air, canopy, and surface temperature. Points represent the difference between treatment and control plots by block (i.e. one 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. Open symbols represent constant wattage control and filled symbols represent feedback control. 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. Error bars represent standard error. See Supplemental Materials (especially Tables S5 and S6) for details.

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 with feedback control 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 with feedback control, 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 and feedback control, 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, in which soil temperature was measured at a depth of 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; Hoeppner & Dukes 2012; Wolkovich *et al.* 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

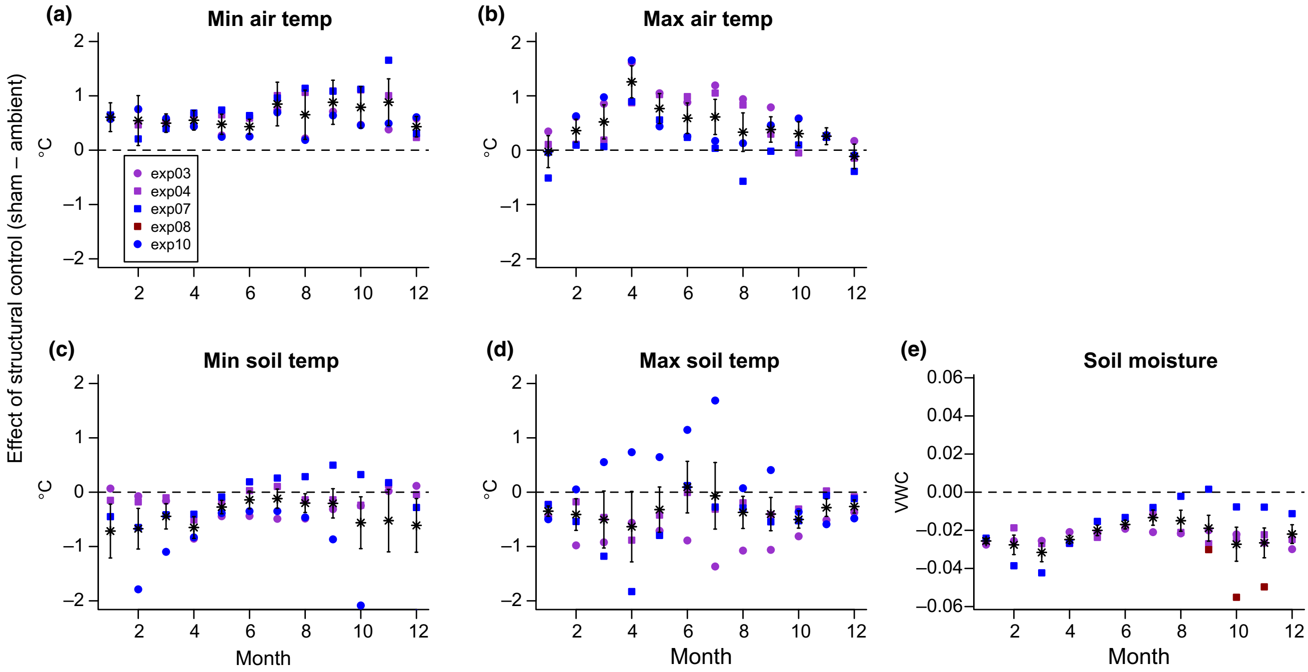

Figure 4 Deviations in measured abiotic variables by month in structural controls compared to ambient controls (i.e. with no control chambers or warming infrastructure in place). Above-ground temperatures (which include includes air, canopy, and surface temperatures) were higher (a,b), whereas soil temperature (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). Colours 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). All studies included used feedback warming control. See Supplemental Materials for details (Tables S8–S13).

ambient controls so that they may separate artefacts due to infrastructure from the effects of experimental warming. Increased use of both structural and ambient controls together would also help answer important questions of how infrastructure effects vary across ecosystem types, and warming designs (Box 1, Fig. 2).

INDIRECT AND FEEDBACK EFFECTS OF CLIMATE CHANGE MANIPULATIONS

Climate change experiments often seek to manipulate temperature or precipitation separately as well as interactively, and yet 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 three experiments that manipulated both temperature and precipitation, and provided daily above-ground and soil temperature data. Across these studies, all of which used infrared heating (two with feedback control and one with constant warming), we found that increasing the amount of added precipitation reduced daily minimum and maximum above-ground temperatures, at rates of 0.01 °C 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 favour latent (e.g. 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 vapour 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 11 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 (Fig. 5, Table S15). The drying effect varied by warming type (−0.80% for infrared vs. −0.33% for forced air, per °C of target warming, Table S16).

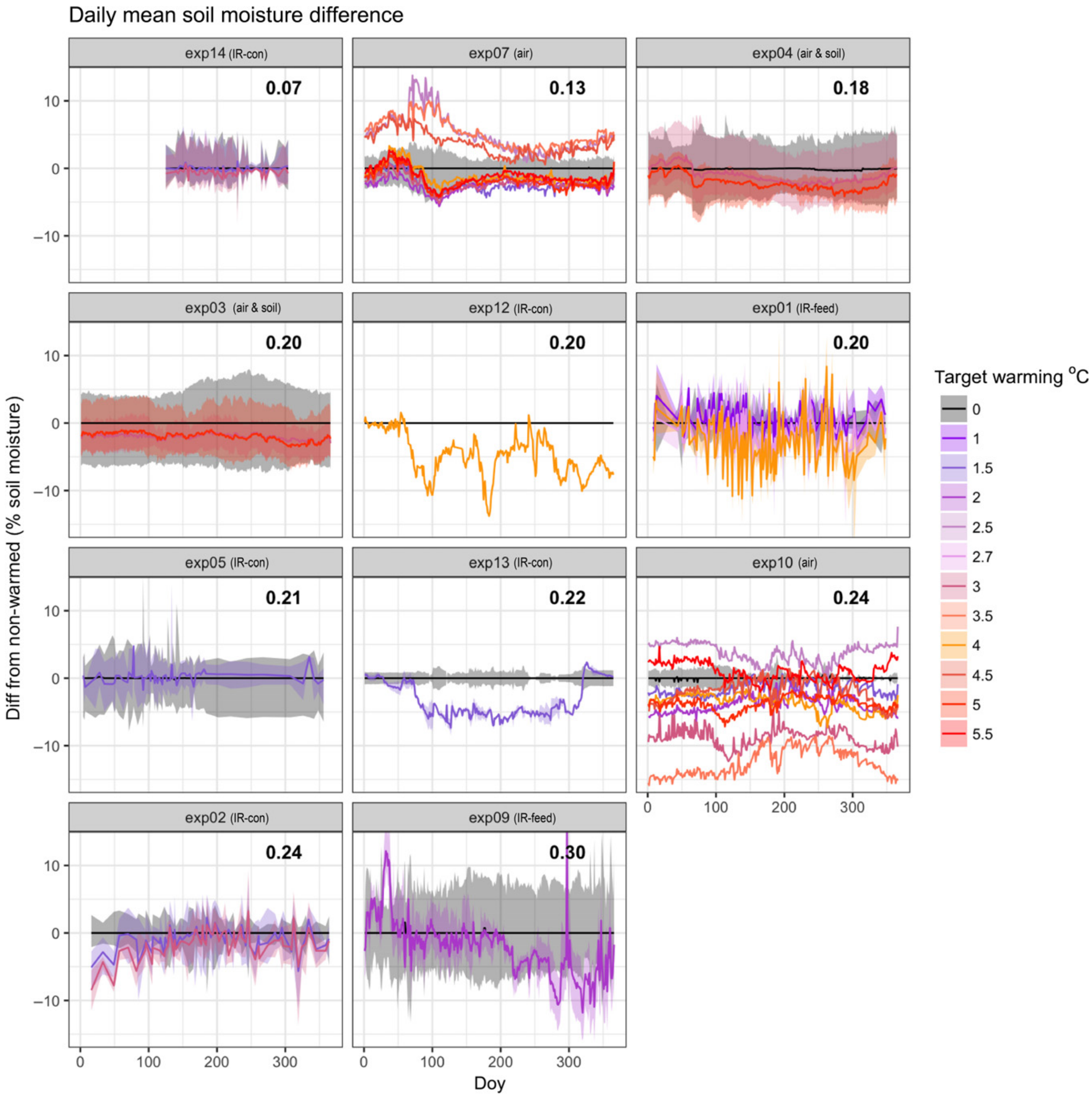

Figure 5 Deviations in daily observed soil moisture, shown for the 11 study sites that continuously monitored soil moisture, excluding data from plots that manipulated precipitation. Black lines represent control plots, and coloured lines represent warming treatments with various target warming levels (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. Heating type is listed in parentheses next to the site number: IR-con = infrared with constant wattage, IR-feed = infrared with feedback control, soil = soil cables, air = forced air.

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.* 2014a,b; Harte *et al.* 2015). These community shifts may affect resource levels, such as moisture, carbon, and nutrient levels in the soil (McDaniel *et al.* 2014a,b; Harte *et al.* 2015) and feed back to affect microclimate (Harte *et al.* 2015).

The presence of these feedback effects is both a strength 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 artefacts 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 – analyse warming and/or precipitation treatments as simple categorical predictors (e.g. as in a two-way ANOVA). Our findings, however, demonstrate a 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. Pelini *et al*. 2014; Marchin *et al*. 2015).

ECOLOGICAL IMPLICATIONS

We have highlighted a suite of factors that complicate interpretation of climate change experiments. These indirect effects are likely to have biological implications for many of the responses studied in warming experiments (e.g. Fig. 6). 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 (Fig. 6b) or exaggerated (Fig. 6c) 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 vs. plot-level climate variables. We fit two separate mixed-effects models, which 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 modelled other differences between studies that may have affected phenology), year (nested within study, which modelled differences due to variability among years, which may have altered phenology), and species (which often vary in their phenology). All random effects were modelled on the intercept only; see Supplemental Materials for details.

We found that phenological sensitivities to temperature estimated from the two modelling 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 Fig. 7), 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

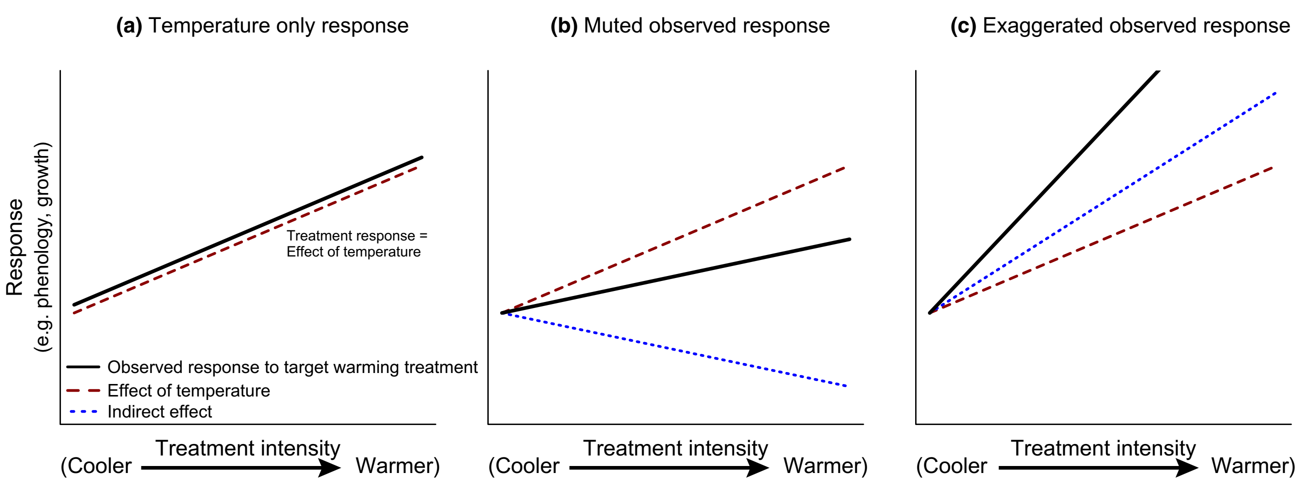

Figure 6 Theoretical biological responses to experimental warming and their interpretation. Direct responses to temperature alone (a) can be easily understood. Complications arise when biological responses are a mix of the direct temperature and indirect non-temperature effects of experimental warming. 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 artefacts 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).

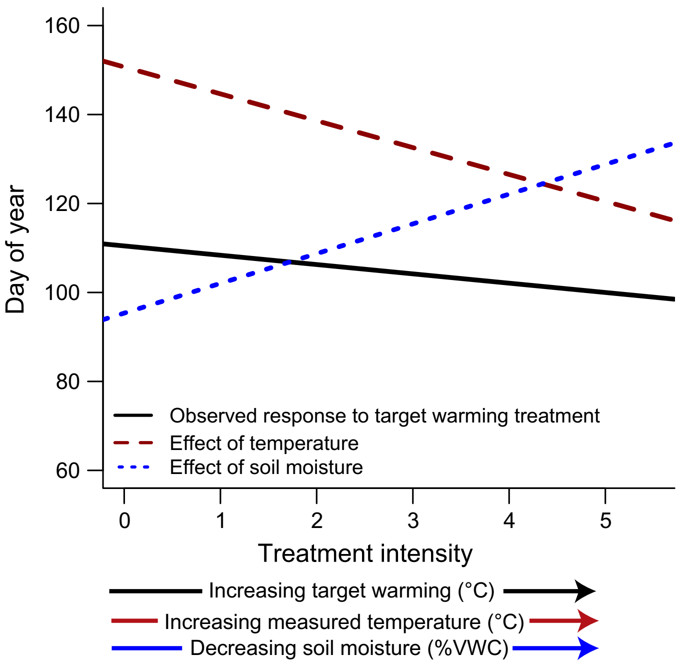

Figure 7 Observed response of budburst day of year to experimental climate change is an example of a muted response: the observed response to increasing treatment intensity (i.e. the coefficient of a model fit with only target [or reported, if there was no explicit target] temperature as the explanatory variable, black line) 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) and soil moisture (dotted blue line, 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 and S18, for additional details.

variables are important drivers of budburst timing (Tables S17 and 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, Fig. 7). This negative effect is expected, if reducing moisture delays budburst (Table S17, Fig. 7), 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 (Fig. 3), 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 (Fig. 6b); 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 vs. 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 artefacts 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. Global minimum temperatures increased more rapidly than maximum temperatures from 1950 to 1980, reducing above-ground daily temperature range (Vose *et al.* 2005; Thorne *et al.* 2016). 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, including indirect effects of treatments, 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. Although surface warming inevitably increases soil water evaporation, it does not necessarily translate to a decrease in soil water content. Precipitation forecasts with climate change are more uncertain than temperature forecasts, as are, consequently, future changes in soil moisture (Cook *et al.* 2018). For example, soil drying 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. Shifts in soil moisture are likely to vary by region, season, vegetation type, and soil depth (Seager *et al.* 2014; Berg *et al.* 2017; Cook *et al.* 2018). 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 quantify soil moisture effects in climate change experiments. The altered light, wind, and herbivory patterns documented under experimental infrastructure (Kennedy 1995; Moise & Henry 2010; Hoeppner & Dukes 2012; Wolkovich *et al.* 2012; Clark *et al.* 2014b) represent other non-temperature effects that may be potential experimental artefacts 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, and we need a fuller understanding of the variable effects across different warming methodologies. We have compiled a database of microclimate data from multiple warming experiments and shown how time, space, experimental artefacts, and indirect effects of treatments may complicate interpretations of these experimental results. The relative importance of each of these factors is likely to vary across warming designs (Box 1), as well as myriad other attributes of sites, making more studies that measure climate similarly and include full infrastructure controls important for progress. 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 2), 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 them 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.

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 data sets; AKE and CRR analysed the data and created the figures; AKE wrote the manuscript.

DATA ACCESSIBILITY STATEMENT

Data available from the Dryad Digital Repository: http://doi.org/10.5061/dryad.p46k773.

REFERENCES

Andresen, L.C., Müller, C., de Dato, G., Dukes, J.S., Emmett, B.A., Estiarte, M. *et al*. (2016). Shifting impacts of climate change: long-term patterns of plant response to elevated CO_2, drought, and warming across ecosystems. In: *Advances in Ecological Research* (ed. Dumbrell, A.J., Kordas, R.L. & Woodward, G.). Elsevier, London, vol. 55, pp. 437–473.

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

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

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

Bokhorst, S., Huiskes, A., Aerts, R., Convey, P., Cooper, E.J., Dalen, L. *et al*. (2013). Variable temperature effects of open top chambers at polar and alpine sites explained by irradiance and snow depth. *Glob. Change Biol*., 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. *et al*. (2016). Temperature response of soil respiration largely unaltered with experimental warming. *Proc. Natl Acad. Sci*., 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., Morin, X., Sonié, L., Collin, C., Fabreguettes, J., Degueldre, D. *et al*. (2012). Climate change might increase the invasion potential of the alien c4 grass setaria parviflora (Poaceae) in the Mediterranean Basin. *Divers. Distrib*., 18, 661–672.

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. *Glob. Change Biol*., 22, 3444–3460.

Clark, J.S., Melillo, J., Mohan, J. & Salk, C. (2014a). The seasonal timing of warming that controls onset of the growing season. *Glob. Change Biol.*, 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. *Funct. Ecol.*, 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. *Proc. Natl Acad. Sci. USA*, 103, 13740–13744.

Cook, B.I., Mankin, J.S. & Anchukaitis, K.J. (2018). Climate change and drought: from past to future. *Curr. Clim. Change Rep.*, 4, 164–179.

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. *Nat. Clim. Chang.*, 3, 52–58.

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

De Boeck, H.J., Vicca, S., Roy, J., Nijs, I., Milcu, A., Kreyling, J. *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. *Integr. Comp. Biol.*, 53, 965–974.

Dukes, J.S. & Mooney, H.A. (1999). Does global change increase the success of biological invaders? *Trends Ecol. Evol.*, 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. *Ecol. Monogr.*, 73, 69–86.

Elmendorf, S.C., Henry, G.H., Hollister, R.D., Björk, R.G., Bjorkman, A.D., Callaghan, T.V. *et al.* (2012). Global assessment of experimental climate warming on tundra vegetation: heterogeneity over space and time. *Ecol. Lett.*, 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 J.*, 7, 31–38.

Ettinger, A. & Wolkovich, E. (2018). MicroClimate from Climate Change Experiments. Knowledge Network for Biocomplexity. https://doi.org/10.5063/F1QV3JQR.

Famiglietti, J., Devereaux, J., Laymon, C., Tsegaye, T., Houser, P., Jackson, T. *et al.* (1999). Ground-based investigation of soilmoisture variability within remote sensing footprints during the southern great plains 1997 (sgp97) hydrology experiment. *Water Resour. Res.*, 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. *J. Ecol.*, 83, 967–977.

Field, C.B., Lobell, D.B., Peters, H.A. & Chiariello, N.R. (2007). Feedbacks of terrestrial ecosystems to climate change. *Annu. Rev. Environ. Resour.*, 32, 1–29.

Franklin, J.F. (1989). Importance and justification of long-term studies in ecology. In: *Long-term Studies in Ecology* (ed. Likens, G.E.). Springer, New York, Berlin, Heidelberg, 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. *J. Clim.*, 27, 511–526.

Giasson, M.A., Ellison, A.M., Bowden, R., Crill, P.M., Davidson, E., Drake, 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. *et al.* (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. *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. *Glob. Ecol. Biogeogr.*, 13, 469–471.

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. *Ecol. Appl.*, 5, 132–150.

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. *Glob. Change Biol.*, 21, 2349–2356.

Helm, B. & Shavit, A. (2017). Dissecting and reconstructing time and space for replicable biological research. In: *Stepping in the Same River Twice: Replication in Biological Research* (eds Shavit, A. & Ellison, A.M.). Yale University Press, New Haven, CT, 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. *Glob. Change Biol.*, 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. *Ecol. Monogr.*, 54, 187–211.

Ibanez, I., Clark, J.S., Dietze, M.C., Feeley, K., Hersh, M., LaDeau, S. *et al.* (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 J.*, 43, 202–206.

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

Kennedy, A. (1995). Temperature effects of passive greenhouse apparatus in high-latitude climate change experiments. *Funct. Ecol.*, 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. *J. Anim. Ecol.*, 84, 785–796.

Kimball, B. (2005). Theory and performance of an infrared heater for ecosystem warming. *Glob. Change Biol.*, 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. *Glob. Change Biol.*, 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. *Glob. Change Biol.*, 21, 3138–3151.

Marion, G., Henry, G., Freckman, D., Johnstone, J., Jones, G., Jones, M. *et al.* (1997). Open-top designs for manipulating field temperature in high-latitude ecosystems. *Glob. Change Biol.*, 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. *Theoret. Appl. Climatol.*, 116, 287–299.

Melillo, J., Frey, S., DeAngelis, K., Werner, W., Bernard, M., Bowles, F. *et al.* (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. *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 Ecol. Evol.*, 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 Phytol.*, 186, 900–910.
Norby, R., Edwards, N., Riggs, J., Abner, C., Wullschleger, S. & Gunderson, C. (1997). Temperature-controlled open-top chambers for global change research. *Global Change Biology*, 3, 259–267.
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. *Glob. Ecol. Biogeogr.*, 15, 395–405.
Parmesan, C. (2006). Ecological and evolutionary responses to recent climate change. *Annu. Rev. Ecol. Evol. Syst.*, 37, 637–669.
Pearson, R.G. & Dawson, T.P. (2004). Bioclimate envelope models: what they detect and what they hide – response to Hampe (2004). *Glob. Ecol. Biogeogr.*, 13, 471–473.
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 Ecol. Evol.*, 2, 534–540.
Pelini, S., Diamond, S., Nichols, L., Stuble, K., Ellison, A.M., Sanders, N. *et al.* (2014). Geographic differences in effects of experimental warming on ant species diversity and community composition. *Ecosphere*, 5, 1–12.
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.
Peterjohn, W.T., Melillo, J.M., Steudler, P.A., Newkirk, K.M., Bowles, F.P. & Aber, J.D. (1994). Responses of trace gas fluxes and N availability to experimentally elevated soil temperatures. *Ecological applications*, 4, 617–625.
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. *Sci. Rep.*, 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. *Int. J. Biometeorol.*, 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, Cambridge.
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. *Nat. Clim. Chang.*, 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.
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.
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 below-ground warming in the B4WarmED (Boreal forest warming at an ecotone in danger) experiment. *Glob. Change Biol.*, 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. *Glob. Change Biol.*, 18, 1108–1116.
Rollinson, C.R. & Kaye, M.W. (2015). Modeling monthly temperature in mountainous ecoregions: importance of spatial scale for ecological research. *Clim. Res.*, 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 Biogeochem. Cycles*, 16, 3–1–3–18.
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. *Nat. Clim. Chang.*, 3, 482.
Seager, R., Neelin, D., Simpson, I., Liu, H., Henderson, N., Shaw, T. *et al.* (2014). Dynamical and thermodynamical causes of large-scale changes in the hydrological cycle over north america in response to global warming. *J. Clim.*, 27, 7921–7948.
Shaver, G.R., Canadell, J., Chapin, F.S., Gurevitch, J., Harte, J., Henry, G. *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. *Ecol. Lett.*, 14, 1191–1200.
Sherry, R.A., Zhou, X., Gu, S., Arnone, J.A., Schimel, D.S., Verburg, P.S. *et al.* (2007). Divergence of reproductive phenology under climate warming. *Proc. Natl Acad. Sci. USA*, 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. *Quatern. Res.*, 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. *et al.* (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. *J. Biogeogr.*, 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. *Environ. Entomol.*, 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. *Geophys. Res. Lett.*, 32.
Thomas, C.D., Cameron, A., Green, R.E., Bakkenes, M., Beaumont, L.J., Collingham, Y.C. *et al.* (2004). Extinction risk from climate change. *Nature*, 427, 145–148.
Thorne, P., Donat, M., Dunn, R., Williams, C., Alexander, L., Caesar, J. *et al.* (2016). Reassessing changes in diurnal temperature range: intercomparison and evaluation of existing global data set estimates. *J. Geophys. Res. Atmos.*, 121, 5138–5158.
Thuiller, W. (2004). Patterns and uncertainties of species' range shifts under climate change. *Glob. Change Biol.*, 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. *Proc. R. Soc. B Biol. Sci.*, 279, 2072–2080.
Vasseur, D.A., DeLong, J.P., Gilbert, B., Greig, H.S., Harley, C.D., McCann, K.S. *et al.* (2014). Increased temperature variation poses a greater risk to species than climate warming. *Proc. R. Soc. Lond. B Biol. Sci.*, 281, 20132612.
Vose, R.S., Easterling, D.R. & Gleason, B. (2005). Maximum and minimum temperature trends for the globe: an update through 2004. *Geophys. Res. Lett.*, 32, e01221.

Williams, J.W. & Jackson, S.T. (2007). Novel climates, no-analog communities, and ecological surprises. *Front. Ecol. Environ.*, 5, 475–482.
Williams, J.W., Jackson, S.T. & Kutzbacht, J.E. (2007). Projected distributions of novel and disappearing climates by 2100 AD. *Proc. Natl Acad. Sci. USA*, 104, 5738–5742.
Wolkovich, E.M., Cook, B.I., Allen, J.M., Crimmins, T.M., Betancourt, J.L., Travers, S.E. *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 Phytol.*, 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. *Proc. Natl Acad. Sci.*, 113, 10589–10594.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Editor, Elizabeth Jeffers
Manuscript received 4 May 2018
First decision made 18 June 2018
Second decision made 30 August 2018
Third decision made 28 November 2018
Manuscript accepted 17 December 2018

Supplemental Materials for: How do climate change experiments alter plot-scale climate?

A.K. Ettinger, I. Chuine, B. Cook, J. Dukes, A.M. Ellison, M.R. Johnston, A.M. Panetta, C. Rollinson, Y. Vitasse, E. Wolkovich

December 19, 2018

Additional methods for database development

For our literature review, we searched Web of Science (ISI) for Topic=(warm* OR temperature*) AND Topic=(plant* AND phenolog*) AND Topic=(experiment* OR manip*). We restricted dates to the time period after the STONE database (i.e., January 2011 through March 2015). This yielded 277 new studies. We removed all passive warming studies from the list, and contacted authors for daily data. Three additional datasets were offered or suggested to us, and in March 2018, we added additional studies found by using the same terms to search the following online databases for additional datasets: dataONE (https://search.dataone.org/), KNB (https://knb.ecoinformatics.org/), and dryad (https://www.datadryad.org/). The resulting database contains daily climate data collected between 1991 and 2015 from 15 North American, European, and Chinese climate change experiments (Table S1, Figure S1).

Details of statistical analyses and results

For all analyses, we used mixed-effects models implemented with the lme4 package in R, version 3.4.2 (Bates et al., 2015; R Core Team, 2017). Mixed-effects models, also called multi-level or hierarchical models, can account for structured data that violate the independence assumption of traditional linear regression (Gelman and Hill, 2007). In our analyses, we use levels/groupings of experimental site, year, and day of year (DOY) to account for this mutual dependence among data points. We use three as a minimum threshold for sample size within each level/grouping for random effects (i.e., there must be at least three units within each level/group for it to be considered as a separate group in our analyses). To test for significance of fixed effects in our models, we use Type II tests for models including only main effects and Type III tests for models including interactions and main effects.

Analysis of effects of warming on daily temperature range

To test how active warming alters daily temperature range (DTR, the difference between maximum and minimum temperatures in a day), we used data from eight (for above-ground) and ten (for soil) studies in the MC3E database that include daily measurements of above-ground (i.e., air, canopy, or surface) and soil temperature maxima and minima. These data were from studies that used infrared warming with feedback control (n=2), infrared with constant control (n=4), forced air with feedback control (n=2), and combined forced air and soil cable with feedback control (n=2). For consistency, we included only structural controls (we therefore excluded exp15, which used only ambient controls), and also excluded data from plots with precipitation treatments. We fit linear mixed-effect models with above-ground DTR, soil DTR, minimum and maximum daily above-ground temperature, and minimum and maximum daily soil temperature as response variables. We included target temperature treatment (or measured temperature, for those studies that did not have explicit target temperatures) as a predictor (i.e., a fixed effect). Random effects were site and study year nested within site (with random slopes and intercepts, Tables S3 & S4).

Analysis of effects of time and space on experimental microclimate

To test how treatment effects vary spatially (i.e., among blocks within a study) and temporally (i.e., among years within a study), we used data from the six studies in the MC3E database that used blocked designs. These studies included infrared (n=5), and soil cable (n=1) warming types. We fit linear mixed-effect models with mean daily soil temperature, minimum daily above-ground temperature, and maximum daily above-ground temperature as response variables (Figure 3 in the main text). For temporal models, we included fixed effects of temperature treatment, year, and their interaction; random effects were site and block nested within site (intercept-only structure, Table S5). For spatial models, we included fixed effects of temperature treatment, block, and their interaction; random effects were site and year nested within site (intercept-only structure, Table S6). Both of these models excluded data from plots with precipitation treatments.

Analysis of effects of infrastructure on experimental microclimate

To test how infrastructure affects microclimate, we compared temperature and soil moisture data from the 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 or forced air chambers with no heat applied) and ambient controls with no infrastructure added. These five studies consisted of soil cables (exp08), forced air (exp07, exp10), and combined soil and forced air (exp03, exp04) warming types (all with feedback control) and occurred at two sites: Duke Forest and Harvard Forest (Farnsworth et al., 1995; Clark et al., 2014; Marchin et al., 2015; Pelini et al., 2011). One additional study, exp15, which utilized forced air, monitored environmental conditions in both ambient and structural controls, but we were only able to obtain data for the ambient controls so it is excluded from this analysis. Note that all studies that employ forced air warming utilize chambers, whereas the other warming types did not utilize chambers.

For experiments using infrared heating, we were unable to directly compare ambient and structural controls because no studies in our database shared data from both control types. Instead, we compared differences between each control type (ambient versus structural) and the measured amount of warming per degree of target warming in experimental plots. We predicted that, if heating infrastructure affects microclimate, the measured amount of warming per degree of target warming should differ by control type. Of the seven infrared sites in our database that measured soil temperature, three used ambient controls (exp06, exp11, exp14) and four used structural controls (exp01, exp09, exp12, exp13). Two studies used feedback control (exp01, exp09) and five studies used constant warming control (exp06, exp11, exp12, exp13, exp14). Among these studies, warming per °C of target warming was 1.01°C for structural controls (SE=0.09, Table S7), but significantly lower for ambient controls (0.41°C of warming per °C of target warming, SE=0.09, Table S7). This trend, based on a limited number of studies and sites, suggests that infrared heating equipment may alter microclimate, likely by shading the soil surface (McDaniel et al., 2014).

For studies for which we were able to directly compare ambient and structural controls, we fit linear mixed-effects models by month for the following response variables: mean daily soil temperature, and minimum and maximum daily air and soil temperature (Farnsworth et al. (1995) only measured mean soil, so there are only four different studies in models for minimum and maximum air and soil), and soil moisture. The predictor was control type (sham or ambient). To allow for both mean differences in temperature and the effect of control type to vary among sites and years, random effects were site and year nested within site, modeled as random slopes and random intercepts. We found that experimental structures altered above-ground and soil temperatures in opposing ways: above-ground temperatures were higher in the structural controls, compared with ambient conditions with no structures installed, whereas soil temperatures were lower in the structural controls compared with ambient soil (Figure 4 in the main text). In addition, soil moisture was lower in structural controls compared with ambient conditions. These general patterns were consistent across the different temperature models we fit (mean, minimum, and maximum soil and air temperatures), although the magnitude varied across months, as well as across studies. We show summaries from models fit to the entire year (Tables S8, S9, S12), as well as summaries from models fit to each month of data, as shown in Figure 4 in the main text (Tables S10, S11, S13).

Analysis of effects of precipitation treatments on above-ground temperature

Of the 15 experiments in the MC3E database, three manipulated precipitation and provided above-ground and soil temperature data (all used infrared warming, with one constant wattage and two feedback control). To examine the effects of precipitation treatment on temperature, we fit linear mixed-effects models to data from these three sites, with temperature (above-ground daily minimum and maximum, and soil minimum and maximum) as the response variables. Predictors were precipitation treatment (a continuous fixed effect, which ranged from 50 to 200% of ambient for these three studies), target warming (a continuous fixed effect, which ranged from 0 to 4 °C for these three studies), and their interaction. To account for methodological and other differences among sites, we included site as a random effect, with year and DOY nested within site to account for the non-independent nature of measurements taken on the same day within sites. We used a random intercept model structure (Table S14).

Analysis of effects of experimental warming on soil moisture

Of the 15 experiments in the MC3E database, 12 measured and reported soil moisture. These studies included infrared with feedback control (n=2), infrared with constant energy output (n=5), forced air with feedback control (n=3), and combined forced air and soil cable with feedback control (n=2). To examine the effects of target warming treatment on soil moisture, we fit linear mixed-effects models to data from these ten sites, excluding plots with precipitation treatments. We first fit a model with soil moisture as the response and a predictor of target warming (a continuous fixed effect, which ranged from 0 to 5.5 °C for these 12 studies). To account for differences among warming types, we included warming type as a fixed effect, grouping all infrared types together and grouping combined forced air/soil cable warming type with forced air studies because of the low sample sizes. We accounted for other differences among studies by including site as a random effect, with year and DOY nested within site to account for the non-independent nature of measurements taken on the same day within sites. We used a random intercept model structure (Table S16).

In addition to testing how experimental warming influenced soil moisture, we also tested how experimental structures influenced soil moisture. We compared the soil moisture measured in structural controls to both ambient controls and warmed plots by fitting a model with categorical fixed effects of "ambient," "structural control," and "warmed." We again included warming type as a fixed effect, and site as a random effect, with DOY nested within site to account for the non-independent nature of measurements taken on the same day within sites, using a random intercept structure (Table S15).

Analysis of budburst phenology

We wanted to investigate how using measured plot-level climate variables, as opposed to target warming, alters estimates of temperature sensitivity in ecology. To do this, we fit two different types of models to data from the five study sites in the MC3E database that recorded above-ground temperature and soil moisture, as well as phenology data (DOY of budburst). These studies included infrared (n=1), forced air (n=2), and combined forced air and soil cable (n=2) warming types. (All of these studies used feedback control). We used only structural controls in the reported analysis, because this is the type of control that all five studies possess (including ambient control plots in the analysis did not qualitatively change the results). We focus on budburst, as this phenological phase was reported most commonly among studies in the MC3E database. For all models, we accounted for non-independence by including species, site, and year nested within site as intercept-only random effects (Table S17). The target warming model included only one explanatory variable (the target amount of warming). We compared this to models with mean annual measured above-ground temperature (offset by subtracting the minimum temperature across all studies and plots, to make model intercepts more similar), mean winter (January-March) soil moisture, and their interaction as explanatory variables. The slope for temperature in the measured climate model can be directly compared to the slope for target warming in the target warming model because the units are the same (change in budburst, in days, per°C).

To determine which specific above-ground temperature variable to include in the measured climate model, we compared Akaike's information criterion (AIC) values of models fit with four different temperature variables (mean annual minimum and maximum temperatures, mean January-March minimum and maximum temperatures). The model with mean annual minimum temperature, mean January-March soil moisture, and their interaction provided the best model fit (lowest AIC, highest explained variation, Table S18), so we discuss and interpret that model in the main text, summarize it in Table S17, and present its coefficients in Figure 7.

References

Bates, D., M. Maechler, B. M. Bolker, and S. Walker. 2015. Fitting linear mixed-effects models using lme4. Journal of Statistical Software 67:1–48.

Bokhorst, S., A. Huiskes, R. Aerts, P. Convey, E. J. Cooper, L. Dalen, B. Erschbamer, J. Gudmundsson, A. Hofgaard, R. D. Hollister, 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.

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

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

Gelman, A., and J. Hill. 2007. Data Analysis Using Regression and Multilevel/Hierarchical Models. Cambridge University Press, New York, NY, USA.

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

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

Pelini, S. L., F. P. Bowles, A. M. Ellison, N. J. Gotelli, N. J. Sanders, and R. R. Dunn. 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.

R Core Team. 2017. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Supplemental Tables

Table S1: **Experimental sites included in the MC3E database**. Experimental sites correspond to the map (Figure S1). We give the study ID, location, source, years of data included, warming type, the type of control used to maintain warming, plot area (in square meters), watts of heating output, target warming treatment (°C; reported warming was used for exp06, exp11, exp12, and for the two lower warming treatments for exp01, which did not have explicit targets for warming treatments), precipitation treatment (proportion of ambient), method of above-ground temperature measurement (with height of measurement, in cm, for air), depth of soil temperature measurement (cm), depth of soil moisture measurement (cm) used in each study, the type(s) of control plots installed (structural controls contain all the warming infrastructure, such as soil cables, but with no heat applied; ambient controls have no infrastructure added), and statistical analysis used in the source listed (i.e., ANOVA with categorical explatory variables for different warming levels versus continuous microclimate explanatory variables). All studies that employ forced air warming utilize chambers, whereas the other warming types did not utilize chambers. Note that some sites may have multiple sources; however, we list only one here. *data collected but not available in the MC3E database.

study	location	source	data years	warming type	warming control	area	watts	warming trtmt	precip trtmt	above-ground temp	soil temp depth	soil moist depth	control type	analysis type
exp01	Waltham, MA, USA	Hoeppner and Dukes 2012	2009-2011	infrared	feedback	4.00	50, 150, 250	1, 2.7, 4	0.5, 1.0, 1.5	canopy	2, 10	30	structural	categorical
exp02	Montpelier, France	Morin et al. 2010	2004	infrared	feedback	1.56	102.4	1.5, 3	0.7, 1.0			15, 30	ambient	categorical
exp03	Duke Forest, NC, USA	Clark et al. 2014	2009-2014	forced air and soil	feedback	17.00		3, 5		air (30)	10	30	both	continuous
exp04	Harvard Forest, MA, USA	Clark et al. 2014	2009-2012	forced air and soil	feedback	17.00		3, 5		air (30)	10	30	both	continuous
exp05	Jasper Ridge Biological Preserve, CA, USA	Cleland et al. 2007	1998-2002	infrared	constant	3.14	80	1.5	1.0, 1.5		15*	15	ambient	categorical
exp06	Rocky Mountain Biological Lab, CO, USA	Dunne et al. 2003	1995-1998	infrared	constant	30.00	22	1.9			12, 25	12, 25	ambient	categorical
exp07	Harvard Forest, MA, USA	Pelini et al. 2011	2010-2015	forced air	feedback	15.70		1.5-5.5		air (22)	2, 6	30	both	continuous
exp08	Harvard Forest, MA, USA	Farnsworth et al. 1995	1993	soil warming	feedback	36.00		5			5		both	categorical
exp09	Stone Valley Forest, PA, USA	Rollinson and Kaye 2012	2009-2010	infrared	feedback	4.00	100	2	1.0, 1.2	surface	3	8	structural	categorical
exp10	Duke Forest, NC, USA	Marchin et al. 2015	2010-2013	forced air	feedback	15.70		1.5-5.5		air (22)	2, 6	30	both	continuous
exp11	Rocky Mountain Biological Lab, CO, USA	Price and Wasser 1998	1991-1994	infrared	constant	30.00	15	1.2			12		ambient	categorical
exp12	Kessler Farm Field Laboratory, OK, USA	Sherry et al. 2007	2003	infrared	constant	6.00	100	4.17	1.0, 2.0	air (14)	7.5, 22.5	15	structural	categorical
exp13	Haibei Alpine Grassland Research Station, China	Suonan et al. 2017	2012-2014	infrared	constant	3.96	303	1.5		air (30)	5, 10	5, 10	structural	categorical
exp14	Cedar Creek, MN, USA	Whittington et al 2015	2009-2011	infrared	constant	7.50	80, 133	1.5, 3		air (10,25)	1, 10	6	ambient	categorical
exp15	Oak Ridge, TN, USA	Gunderson et al 2015	2003-2005	forced air	feedback	9.42		2, 4		air		10, 20*	both	categorical

Table S2: **Summary of warming treatments, by warming type,** for studies included in the MC3E database. An additional common type of experimental warming, passive open-top chambers (OTC), is included for comparison. Summaries of the target warming treatments (°C) and measured warming for air temperature, soil surface temperature, and soil temperature (°C) are given. Measured warming shown here is for warming treatments only (precipitation treatments are excluded), and is the difference between mean annual temperature (MAT) of control plots and MAT of each treatment level within a study. Mean difference (with standard error) and the range of differences (minimum to maximum differences in MAT) are shown. n is the number of studies of each warming type in the MC3E database. *Passive OTC data are from Bokhorst et al. (2013).

temperature type	forced air	force air, soil cables	infrared	soil cables	passive otc
target (min-max)	3.5 (1.5-5.5)	4 (3-5)	2.2 (1-4)	5	
air, mean (se)	3.22 (0.12)	1.9 (0.24)	1.08 (0.16)		0.8 (0.1)
air, range	0.83-5.2	-0.05-3.59	0.42-1.83		0.5-1.3
surface, mean (se)			1.72 (0.1)		0.9 (0.1)
surface, range			1.52-1.87		0.4-1.4
soil, mean (se)	1.29 (0.09)	3.01 (0.34)	1.73 (0.2)	5.04	0.8 (0.3)
soil, range	-0.93-3.46	0.02-5.08	-0.06-4.31		-0.1-3.9
n	3	2	9	1	0*

Table S3: **Summary of linear mixed-effects models of how target warming treatment affects above-ground daily temperature range (DTR) and mimumim and maximum temperatures in climate change experiments.** We excluded data from plots with precipitation treatments from these analyses, which included data from sites using infrared with feedback control, infrared with constant control, forced air, and combined forced air and soil warming. Estimates (est.) are the intercept and coefficient for target warming from the model; se is the standard error for these estimates. The effect of target warming on observed warming was significant based on Type II Wald χ^2 tests of fixed effects for above-ground minimum temperature (χ^2=46.58, df=1, p<0.001) and above-ground maximum temperature (χ^2=5.74, df=1, p=0.02), but not for above-ground daily DTR (χ^2=1.38, df=1, p=0.24). Random effects were site (n=8), with a random slope and intercept structure (listed for each model), and year nested within site (n=32 year-site combinations, not shown). Total number of observations=169,797, and units are °C for all three models.

	above-ground DTR		above-ground min temp.		above-ground max temp.	
predictor	est.	se	est.	se	est.	se
intercept	14.74	1.58	6.5	0.89	21.46	1.97
target warming effect	-0.37	0.31	0.81	0.12	0.48	0.2
site random effects	int	target	int	target	int	target
exp01	13.65	0.06	6.48	0.8	19.94	0.85
exp03	12.65	-0.13	7.99	0.49	20.95	0.38
exp04	9.5	0.01	6.36	0.49	15.68	0.49
exp07	10.28	0.37	4.61	0.72	14.04	1.06
exp09	17.37	-0.61	4.72	1.13	21.28	0.56
exp10	12	0.23	8.58	0.73	21.44	0.97
exp12	22.63	-2.57	4.85	1.43	29.65	-0.82
exp14	19.84	-0.3	8.39	0.68	28.68	0.37

Table S4: **Summary of linear mixed-effects models of how target warming treatment affects soil daily temperature range (DTR), mimumim, and maximum temperatures in climate change experiments.** We excluded data from plots with precipitation treatments from these analyses, which included data from sites using infrared with feedback control, infrared with constant control, forced air, and combined forced air and soil warming. Estimates (est.) are the intercept and coefficient for target warming from the model; se is the standard error for these estimates. The effect of target warming on observed warming was significant based on Type II Wald χ^2 tests of fixed effects for soil minimum temperature (χ^2=51.59, df=1, p<0.001) and soil maximum temperature (χ^2=66.33, df=1, p<0.001), but not for soil DTR (χ^2=0.08, df=1, p=0.78). Random effects were site (n=10), with a random slope and intercept structure (listed for each model), and year nested within site (n=41 year-site combinations). Total number of observations=168,767, and units are °C for all three models.

	soil DTR		soil min temp.		soil max temp.	
predictor	est.	se	est.	se	est.	se
(Intercept)	4.13	0.47	10.38	1.07	14.44	1.37
target	0.02	0.07	0.72	0.1	0.75	0.09
site random effects	int	target	int	target	int	target
exp01	6.45	0.24	10.29	0.79	16.29	0.96
exp03	2.83	-0.01	12.35	0.74	15.24	0.74
exp04	2.13	0.06	9.17	0.87	11.12	0.92
exp06	4.19	0.02	7.97	0.68	12.02	0.76
exp07	2.72	0.05	7.28	0.41	10.27	0.51
exp09	3.62	-0.27	9.48	1.16	13.07	0.9
exp10	5.14	0.06	12.35	0.43	17.72	0.51
exp11	3.97	0.02	6.12	0.66	9.92	0.78
exp12	4.05	0.24	12.16	0.69	15.53	0.84
exp14	6.15	-0.22	16.66	0.76	23.17	0.55

Table S5: **Analysis of variance table for temporal linear mixed-effects models** of effect of target warming (temp. treatment), year and their interaction on daily mean soil temperature, minimum above-ground temperature, and maximum above-ground temperature, fit by maximum likelihood. Warming types included soil cables with feedback control (exp08), infrared with feedback control (exp01, exp09), and infrared with constant warming (exp12, exp13, exp14). See Figure 3 in the main text. A significant interaction indicates that warming treatment differs across years in a study. We list degrees of freedom (which are identical across response variables), test statistics, and p-values for Type III Wald χ^2 tests of fixed effects in the models. For all models, random effects were site (n=6 for soil temperature model, n=5 for above-ground temperature models) and block nested within site (intercept-only structure; n=52 for soil, n=55 for above-ground); total number of observations=81,031 for soil and 80,987 for above-ground; units are °C.

		mean soil temp.		min above-ground temp.		max above-ground temp.	
predictor	df	χ^2	p	χ^2	p	χ^2	p
intercept	1	83.80	<0.001	84.59	<0.001	255.91	<0.001
temp. treatment	1	1666.73	<0.001	555.56	<0.001	157.46	<0.001
year	5	442.37	<0.001	75.39	<0.001	274.82	<0.001
temp. treatment*year	5	280.91	<0.001	156.17	<0.001	143.04	<0.001

Table S6: **Analysis of variance table for spatial linear mixed-effects models** of effect of target warming (temp. treatment), block, and their interaction on daily mean soil temperature, minimum above-ground temperature, and maximum above-ground temperature, fit by maximum likelihood. See Figure 3 in the main text. A significant interaction indicates that warming treatment differs across blocks witin a study. We list degrees of freedom (which are identical for all models), test statistics, and p-values for Type III Wald χ^2 tests of fixed effects in the models. For all models, random effects were site (n=6 for soil temperature model, n=5 for above-ground temperature models) and year nested within site (intercept-only structure; n=16 for soil, n=13 for above-ground); total number of observations=81,031 for soil, 80,987 for above-ground.

		mean soil temp.		min above-ground temp.		max above-ground temp.	
predictor	df	χ^2	p	χ^2	p	χ^2	p
intercept	1	368.31	<0.001	82.58	<0.001	623.80	<0.001
temp. treatment	1	593.81	<0.001	257.56	<0.001	118.96	<0.001
block	39	415.60	<0.001	118.67	<0.001	217.86	<0.001
temp. treatment*block	39	116.78	<0.001	163.35	<0.001	254.68	<0.001

Table S7: **Comparison of estimated warming (Tdiff) attained in infrared studies, per degree of target warming, with different types of controls**, structural versus ambient. (For three studies [exp06, exp11, exp12] there was no explicit target temperature so reported warming was used.) Two studies used feedback control (exp01, exp09); five studies used constant warming control (exp06, exp11, exp12, exp13, exp14). Tdiff was calculated by substracting mean annual temperature in control plots from mean annual temperature each treatment level among controls, and in seven infrared studies that measured above-ground warming (including canopy, surface, and air temperature). We divided Tdiff by the target warming for the treatment level, to standardize across studies. We then fit a mixed-effects model with Tdiff as the response, control type (structural or ambient) as the predictor, and random effects of site (n=7) and year nested within site (24 year-site combinations), with an intercept-only structure. We list the estimated Tdiff and its standard error (se) for each control type. Tdiff differed significantly across the two control types, based on Type II Wald χ^2 tests of the fixed effects (χ^2=19.71 , df=1, $p<0.001$). Total number of observations was 37.

	Tdiff (per °C of target)	se
structural controls	1.01	0.09
ambient controls	0.41	0.09

Table S8: **Summaries of linear mixed-effects models comparing effects of ambient versus structural controls on daily mean, minimum, and maximum soil temperature in climate change experiments across the year**. Warming types included soil warming (n=1: exp08), forced air (n=2: exp07, exp10), and combined soil and forced air (n=2: exp03, exp04), all with feedback control. Estimates (est.) are the intercept (representing ambient controls) and coefficient (representing structure effects) from the models; se is the standard error for these estimates. For these annual models, differences between control types were significant based on Type II Wald χ^2 tests of fixed effects for mean soil temperature (χ^2=5.53, df=1, p=0.02) and minimum soil temperature (χ^2=3.87, df=1, p=0.05), but not for maximum soil temperature (χ^2=2.08, df=1, p=0.15). For all models, random effects of site (n=5 for mean model, n=4 for min and max models) and year nested within site (n=21 for mean model, n=20 for min and max models) were fit with a random slope and intercept structure; total number of observations=48,860 for the mean model and 44,530 for the min and max models; units are °C.

	mean soil temp.		min soil temp.		max soil temp.	
predictor	est.	se	est.	se	est.	se
intercept	11.89	1.42	10.81	1.48	13.92	1.61
structure effect	-0.57	0.24	-0.63	0.32	-0.54	0.38

Table S9: **Summaries of linear mixed-effects models comparing effects of ambient versus structural controls on daily minimum and maximum air temperature in climate change experiments, across the year.** Warming types included forced air (n=2: exp07, exp10), and combined soil and forced air (n=2: exp03, exp04), all with feedback control. Estimates (est.) are the intercept (representing ambient controls) and coefficient (representing structure effects) from the models; se is the standard error for these estimates. For these annual models, differences between control types were not significant based on Type II Wald χ^2 tests of fixed effects for minimum air temperature (χ^2=1.07, df=1, p=0.30), nor for maximum air temperature (χ^2=0.01, df=1, p=0.91). For both models, random effects of site (n=4) and year nested within site (n=20) were fit with a random slope and intercept structure; total number of observations=44,085; units are °C.

	min air temp.		max air temp.	
predictor	est.	se	est.	se
intercept	6.29	1.51	17.74	1.81
structure effect	0.36	0.35	0.02	0.21

Table S10: **Summaries of linear mixed-effects models, fit to each month of data, comparing effects of ambient versus structural controls on daily mean, minimum, and maximum soil temperature, fit to each month separately**, consistent with Figure 4 in the main text. Warming types included soil cables (exp08), forced air (exp07, exp10), and combined soil and forced air (exp03, exp04), all with feedback control. Estimates (est.) are the intercept (representing ambient controls) and coefficient for structural effects from the models; se is the standard error for these estimates. We list test statistics, and p-values for Type II Wald χ^2 tests of fixed effects (df=1 for all tests). Random effects of site (n=5 for all mean soil temperature models; n=4 for all min and max soil temperature models) and year nested within site (n=19 or 20 year-site combinations for all mean soil temperature models; n=18 or 19 for all min and max soil temperature models) were fit with a random slope and intercept structure; total number of observations ranged from 3,814 to 4,186; units are °C.

		mean soil temp.				min soil temp.				max soil temp.			
mon	predictor	est.	se	χ^2	p	est.	se	χ^2	p	est.	se	χ^2	p
1	intercept	2.66	1.25	3.63	0.057	2.34	1.21	2.09	0.149	3.92	1.65	13.71	<0.001
	structure effect	-0.45	0.23			-0.72	0.50			-0.35	0.09		
2	intercept	2.86	1.44	13.06	<0.001	2.58	1.26	3.24	0.072	4.66	1.92	1.99	0.158
	structure effect	-0.44	0.12			-0.67	0.37			-0.41	0.29		
3	intercept	5.24	1.78	6.44	0.011	4.66	1.58	3.64	0.056	7.75	2.04	0.92	0.337
	structure effect	-0.44	0.17			-0.44	0.23			-0.50	0.52		
4	intercept	9.98	1.85	8.53	0.003	8.93	1.98	10.52	0.001	13.24	1.80	0.96	0.327
	structure effect	-0.67	0.23			-0.65	0.20			-0.63	0.65		
5	intercept	14.92	1.37	3.85	0.05	13.74	1.54	4.91	0.027	17.54	1.41	0.59	0.441
	structure effect	-0.31	0.16			-0.27	0.12			-0.32	0.42		
6	intercept	18.29	1.58	0	0.972	17.43	1.57	0.76	0.383	20.98	1.78	0.04	0.844
	structure effect	-0.01	0.20			-0.14	0.16			0.09	0.47		
7	intercept	21.07	1.33	0.06	0.815	19.97	1.34	0.45	0.501	23.76	1.46	0.01	0.914
	structure effect	-0.07	0.28			-0.12	0.18			-0.07	0.61		
8	intercept	20.93	1.20	2.56	0.11	19.59	1.29	1.35	0.244	23.23	1.42	1.58	0.209
	structure effect	-0.26	0.16			-0.20	0.17			-0.37	0.30		
9	intercept	18.23	1.24	10.15	0.001	16.94	1.36	0.58	0.445	20.54	1.43	1.74	0.188
	structure effect	-0.36	0.11			-0.21	0.27			-0.40	0.31		
10	intercept	13.03	1.22	10.48	0.001	12.26	1.24	1.39	0.239	15.42	1.39	10.02	0.002
	structure effect	-0.42	0.13			-0.56	0.48			-0.50	0.16		
11	intercept	8.27	1.13	1.87	0.172	7.34	1.23	0.83	0.363	10.11	1.43	3.16	0.075
	structure effect	-0.33	0.24			-0.52	0.57			-0.28	0.16		
12	intercept	5.03	1.21	2.8	0.094	4.38	1.24	1.53	0.215	6.40	1.53	4.83	0.028
	structure effect	-0.40	0.24			-0.61	0.49			-0.26	0.12		

Table S11: **Summaries of linear mixed-effects models, fit to each month comparing effects of ambient versus structural controls on daily minimum and maximum above-ground temperature, fit to each month separately**, consistent with Figure 4 in the main text. Warming types included forced air (exp07, exp10) and combined soil and forced air (exp03, exp04), all with feedback control. Estimates (est.) are the intercept (representing ambient controls) and coefficient for structural effects from the models; se is the standard error for these estimates. We list test statistics, and p-values for Type II Wald χ^2 tests of fixed effects (df=1 for all tests). Random effects of site (n=4 for both models) and year nested within site (n=18 year-site combinations for both models) were fit with a random slope and intercept structure; total number of observations was 3,726; units are °C.

		min air temp.				max air temp.			
mon	predictor	est.	se	χ^2	p	est.	se	χ^2	p
1	intercept	-5.49	1.78	5.27	0.022	5.09	2.60	0.01	0.927
	structure effect	0.61	0.26			-0.03	0.29		
2	intercept	-3.92	1.83	1.41	0.235	7.10	3.03	2.93	0.087
	structure effect	0.55	0.46			0.36	0.21		
3	intercept	-0.08	1.55	8.59	0.003	12.60	2.41	2.75	0.097
	structure effect	0.50	0.17			0.52	0.31		
4	intercept	5.28	1.80	9.33	0.002	19.27	1.92	18.31	$<$0.001
	structure effect	0.55	0.18			1.26	0.29		
5	intercept	11.62	1.46	6.56	0.01	23.49	1.03	7.75	0.005
	structure effect	0.48	0.19			0.77	0.28		
6	intercept	15.45	1.47	10.13	0.001	26.32	1.82	4.4	0.036
	structure effect	0.43	0.14			0.59	0.28		
7	intercept	17.90	1.26	4.47	0.035	28.94	1.25	3.58	0.059
	structure effect	0.85	0.40			0.61	0.32		
8	intercept	17.07	1.43	2.07	0.15	27.39	1.15	0.87	0.35
	structure effect	0.65	0.45			0.33	0.35		
9	intercept	13.34	1.39	4.71	0.03	23.72	1.47	2.66	0.103
	structure effect	0.88	0.41			0.38	0.23		
10	intercept	7.26	1.26	4.27	0.039	17.29	1.70	1.89	0.169
	structure effect	0.79	0.38			0.30	0.22		
11	intercept	1.21	1.25	4.23	0.04	12.79	1.83	2.76	0.097
	structure effect	0.88	0.43			0.26	0.15		
12	intercept	-2.83	1.48	5.29	0.021	7.56	2.38	0.26	0.61
	structure effect	0.43	0.19			-0.11	0.23		

Table S12: **Summary of a linear mixed-effects model comparing effects of ambient versus structural controls on daily soil moisture (% volumetric water content, VWC) in climate change experiments across the year.** Warming types included soil cables (exp08), forced air (exp07, exp10), and combined soil and forced air (exp03, exp04), all with feedback control. Estimates (est.) are the intercept (representing ambient controls) and coefficient for structure effects from the models; se is the standard error for these estimates. For this annual model, the difference between control types was significant based on Type II Wald χ^2 tests of fixed effects (χ^2=89.95, df=1, p<0.001). Random effects of site (n=5) and year nested within site (n=21 year-site combinations) were fit with a random slope and intercept structure; total number of observations=44,468.

	soil moisture (vwc)	
predictor	est.	se
intercept	21.20	1.86
structure effect	-2.43	0.26

Table S13: **Summaries of linear mixed-effects models, fit to each month comparing effects of ambient versus structural controls on soil moisture (% VWC), fit to each month separately**, consistent with Figure 4 in the main text. Warming types included forced air (exp07, exp10) and combined soil and forced air (exp03, exp04), all with feedback control. Estimates (est.) are the intercept (representing ambient controls) and coefficient for structural effects from the models; se is the standard error for these estimates. We list test statistics, and p-values for Type II Wald χ^2 tests of fixed effects; df=1 for all models. Random effects of site (n=4) and year nested within site (n=18 year-site combinations) were fit with a random slope and intercept structure; total number of observations was 3,829.

mon	predictor	est.	se	χ^2	p
1	intercept	22.58	3.23	59.24	<0.001
	structure effect	-2.77	0.36		
2	intercept	22.10	3.24	16.78	<0.001
	structure effect	-2.54	0.62		
3	intercept	23.58	2.43	8.3	0.004
	structure effect	-2.48	0.86		
4	intercept	22.54	2.15	9.24	0.0024
	structure effect	-2.06	0.68		
5	intercept	21.08	2.31	40.17	<0.001
	structure effect	-2.20	0.35		
6	intercept	18.44	1.37	30.78	<0.001
	structure effect	-2.12	0.38		
7	intercept	17.60	2.18	20.22	<0.001
	structure effect	-2.38	0.53		
8	intercept	16.59	1.90	12.95	<0.001
	structure effect	-2.09	0.58		
9	intercept	15.99	1.54	13.2	<0.001
	structure effect	-1.79	0.49		
10	intercept	20.15	1.93	20.9	<0.001
	structure effect	-2.27	0.50		
11	intercept	21.18	1.77	21.9	<0.001
	structure effect	-2.70	0.58		
12	intercept	22.74	2.83	15.64	<0.001
	structure effect	-2.88	0.73		

Table S14: **Summary of a linear mixed-effects models of how precipitation treatment affects temperature in climate change experiments.** We include data from all studies that manipulated precipitation and measured daily above-ground temperature and/or soil temperature (exp01, exp09, exp12). All three studies used infrared heating; two used feedback control and one used constant warming. Estimates (est.) are the intercept and coefficients for precipitation (measured a percentage of ambient conditions) and warming treatments (target or reported warming), as well as their interaction, from the model; se is the standard error for these estimates; p-values represent significance tests for Type III Wald χ^2 tests. Random effects were site (n=3), year of study (n=9), and day of year (DOY) nested within year (n=2,747), with a random intercept structure. Total number of observations was 76,482 for all models.

response	predictors	est.	se	χ^2	df	p
min above-ground temp.	intercept	6.68	0.75	78.59	1	<0.001
	preciptreat	-0.01	0.00	930.79	1	<0.001
	warmtreat	0.74	0.01	4391.28	1	<0.001
	precip*warm	0.00	0.00	62.73	1	<0.001
max above-ground temp.	intercept	23.80	1.09	477.23	1	<0.001
	preciptreat	-0.02	0.00	2384.33	1	<0.001
	warmtreat	0.78	0.02	1819.10	1	<0.001
	precip*warm	-0.00	0.00	31.73	1	<0.001
min soil temp.	intercept	12.15	0.62	378.95	1	<0.001
	preciptreat	-0.01	0.00	2280.12	1	<0.001
	warmtreat	0.71	0.01	5112.44	1	<0.001
	precip*warm	0.00	0.00	50.49	1	<0.001
max soil temp.	intercept	18.27	0.97	356.27	1	<0.001
	preciptreat	-0.01	0.00	1381.13	1	<0.001
	warmtreat	1.26	0.02	5696.28	1	<0.001
	precip*warm	-0.00	0.00	720.76	1	<0.001

Table S15: **Summary of a linear mixed-effects model comparing soil moisture (% VWC)** in climate change experiments with different types of active warming and in experimentally warmed plots with two different control types, structural and ambient controls. We excluded data from plots with precipitation treatments from this analysis. Warming types included infrared with feedback control (exp01, exp09), infrared with constant warming (exp05, exp12, exp13, exp14), forced air (exp07, exp10), and combined soil and forced air (exp03, exp04), all with feedback control. All infrared types were pooled, and forced air plots were pooled with combined forced air and warming. Estimates (est.) are the intercept (representing ambient controls in infrared plots) and coefficient (i.e. differences between the ambient for structural controls of different warming types) for structural controls and warmed plots (pooled across all target warming levels); se is the standard error for these estimates. There were significant differences among control types (structural versus ambient), based on Type II Wald χ^2 tests of the fixed effect (χ^2=7962.77, df=2, p<0.001). There was also a significant interaction between control type and warming type (forced air versus infrared; χ^2=718.45, df=2, p<0.001). Random effects were site (n=11), year nested within site (n=44 site-year combinations), and DOY nested within year (8,894 DOY-year-site combinations) with a random intercept structure. Total number of observations was 135,987.

	est.	se
intercept (infrared)	19.35	1.98
forced air	0.42	3.24
structure effect (infrared)	1.48	0.16
warmed effect (infrared)	-0.46	0.16
forced air*structure effect	-3.69	0.17
forced air*warmed effect	-2.41	0.16

Table S16: **Summary of a linear mixed-effects model of how target warming treatment affects soil moisture (% VWC) in climate change experiments with different types of active warming.** We excluded data from plots with precipitation treatments from this analysis. Estimates (est.) are the intercept (representing infrared sites) and coefficients from the model; se is the standard error for these estimates. Coefficients include: warming type (i.e., the difference in intercepts for forced air compared with infrared; the intercept for forced air types=20.59 + (-2.50)=18.19), target warming (a continuous predictor), and the interaction between warming type and target warming treatment (i.e., the change in the effect of target warming for forced air compared with infrared; the effect of target warming for forced air types=-0.81 + 0.48=-0.33). The effects of warming type was not significant (χ^2=0.30, df=1, p=0.58); target warming was significant, based on Type II Wald χ^2 tests of the fixed effects (χ^2=4271.78, df=1, p<0.001), as was the interaction of warming type and target warming (χ^2=591.70, df=1, p<0.001). Random effects were site (n=11), year of study nested within site (44 year-site combinations), and DOY nested within year (n=8,894 DOY-year-site combinations), with a random intercept structure. Total number of observations was 135,987.

	est.	se
intercept (infrared)	20.59	1.60
forced air	-2.40	2.84
target (infrared)	-0.81	0.02
target*forced air	-0.33	0.02

Table S17: **Comparison of linear mixed-effects models for budburst day of year** that contain either target warming treatment only as a fixed effect or measured mean annual above-ground temperature (°C), mean soil moisture from January through March (in % VWC), and their interaction as fixed effects. Estimates (est.) are the intercept and coefficients from the models; se is the standard error for these estimates. Both models include random effects of site (n=5), year of study nested within site (n=13 year-site combinations), and plant species (n=54), each with a random intercept structure. Total number of observations was 12,549. Analysis includes data from all studies that monitored budburst, and measured soil moisture and above-ground temperature. These studies included warming types of infrared with feedback control (exp01), forced air with feedback control (exp07, exp10), and combined forced air with soil warming and feedback control (exp03, exp04).

model	predictor	est.	se
target warming	intercept	110.14	6.78
	target warming treat	-1.91	0.09
tmean*soilmois	intercept	151.02	4.67
	tmean	-6	0.38
	soilmois	-1.51	0.13
	tmean*soilmois	0.16	0.02

Table S18: **Comparison of budburst model fits** from four models with different fixed effects: 1) target warming only, 2) measured mean annual above-ground temperature (tmean) only, 3) tmean and mean soil moisture from January through March (soilmois), and 4) tmean, soilmois, and their interaction. We additionally compared models with mean annual maximum, mean annual minimum, and seasonal temperature variables; we present only tmean here because this variable provided the best model fit (i.e. lowest AIC).

model	df	AIC	ΔAIC
tmean*soilmois	8	86982.26	0.00
tmean+soilmois	7	87074.83	92.57
target warming	6	87102.23	119.97
tmean	6	87118.44	136.18

Supplemental Figures

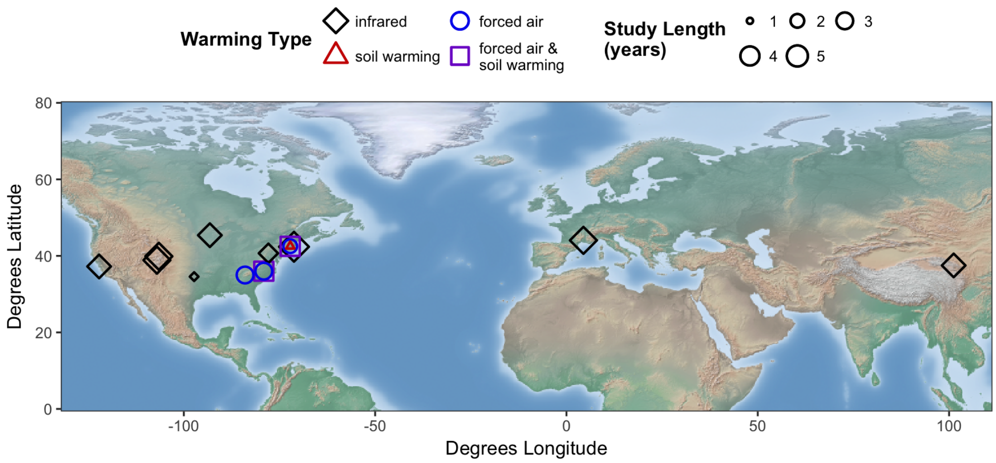

Figure S1: **Climate data from 15 climate change experiments in North America, Europe, and China are included in the MC3E database and analyzed here.** See Table S1 for details.

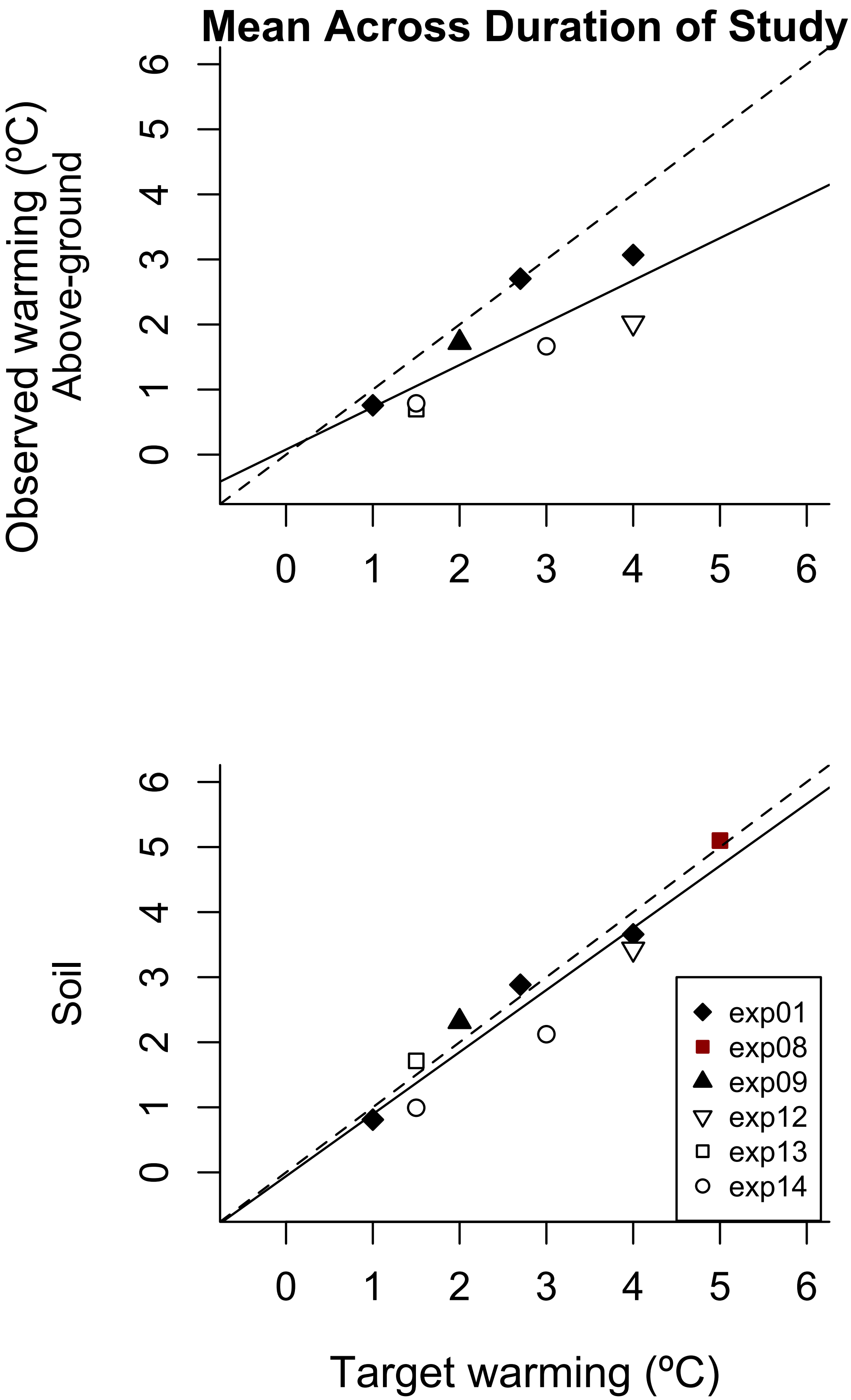

Figure S2: **Mean observed warming, for above-ground and soil temperatures,** excluding data from plots that manipulated precipitation. Above-ground temperature includes air, canopy, and surface temperature. Points represent the difference between treatment and control plots averaged across all plots within a treatment and study, over the duration of the study. 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). Colors vary by heating type: black represents infrared; red represents soil warming cables; open symbols represent constant warming and filled symbols represent feedback control. Compare to Figure 3 in the main text.

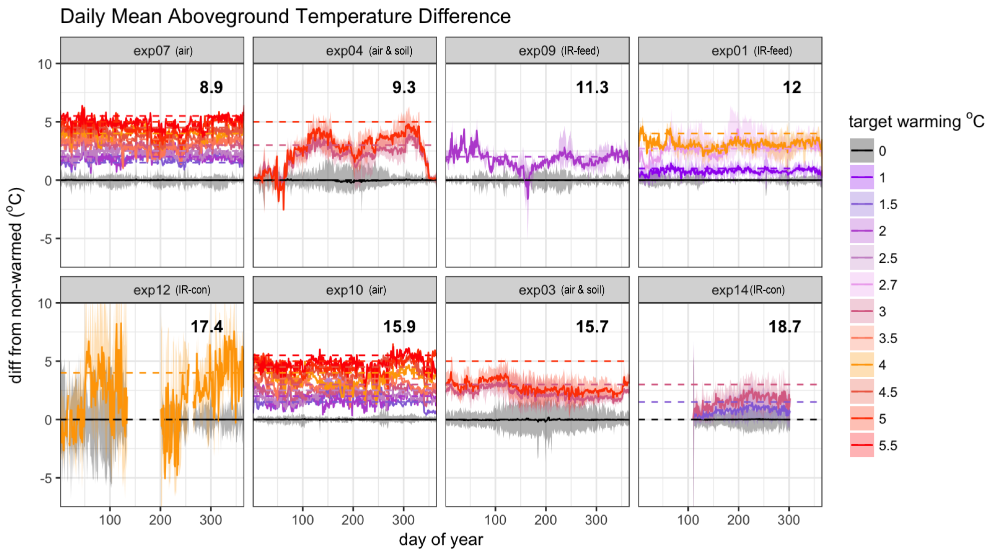

Figure S3: **Deviations in daily observed warming from mean control above-ground temperature for 8 study sites,** excluding data from plots that manipulated precipitation. We show above-ground temperature, which includes studies that measured surface, canopy, and air temperature (see Table S1). Solid 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. Experimental sites are ordered by low to high mean annual temperature (shown in the upper right corner of each panel). Heating type is listed in parentheses next to the site number: IR-con=infrared with constant wattage, IR-feed=infrared with feedback control, soil=soil cables, air=forced air.