

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 12, 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 included and 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, 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 $^{\circ}$ 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 explanatory 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	Hoepfner 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 ($^{\circ}\text{C}$) and measured warming for air temperature, soil surface temperature, and soil temperature ($^{\circ}\text{C}$) 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 minimum 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 ($\chi^2=46.58$, $df=1$, $p<0.001$) and above-ground maximum temperature ($\chi^2=5.74$, $df=1$, $p=0.02$), but not for above-ground daily DTR ($\chi^2=1.38$, $df=1$, $p=0.24$). 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 $^{\circ}\text{C}$ 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), minimum, 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 ($\chi^2=51.59$, $df=1$, $p<0.001$) and soil maximum temperature ($\chi^2=66.33$, $df=1$, $p<0.001$), but not for soil DTR ($\chi^2=0.08$, $df=1$, $p=0.78$). 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 within 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.

predictor	df	mean soil temp.		min above-ground temp.		max above-ground temp.	
		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 subtracting 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 ($\chi^2=19.71$, 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 ($\chi^2=5.53$, df=1, $p=0.02$) and minimum soil temperature ($\chi^2=3.87$, df=1, $p=0.05$), but not for maximum soil temperature ($\chi^2=2.08$, 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 ($\chi^2=1.07$, df=1, p=0.30), nor for maximum air temperature ($\chi^2=0.01$, 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.

mon	predictor	mean soil temp.				min soil temp.				max soil temp.			
		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.

mon	predictor	min air temp.				max air temp.			
		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 ($\chi^2=89.95$, $df=1$, $p<0.001$). 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
	preciptrat	-0.01	0.00	930.79	1	<0.001
	warmtrat	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
	preciptrat	-0.02	0.00	2384.33	1	<0.001
	warmtrat	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
	preciptrat	-0.01	0.00	2280.12	1	<0.001
	warmtrat	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
	preciptrat	-0.01	0.00	1381.13	1	<0.001
	warmtrat	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 ($\chi^2=7962.77$, $df=2$, $p<0.001$). There was also a significant interaction between control type and warming type (forced air versus infrared; $\chi^2=718.45$, $df=2$, $p<0.001$). 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 ($\chi^2=0.30$, $df=1$, $p=0.58$); target warming was significant, based on Type II Wald χ^2 tests of the fixed effects ($\chi^2=4271.78$, $df=1$, $p<0.001$), as was the interaction of warming type and target warming ($\chi^2=591.70$, $df=1$, $p<0.001$). 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 ($^{\circ}\text{C}$), 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

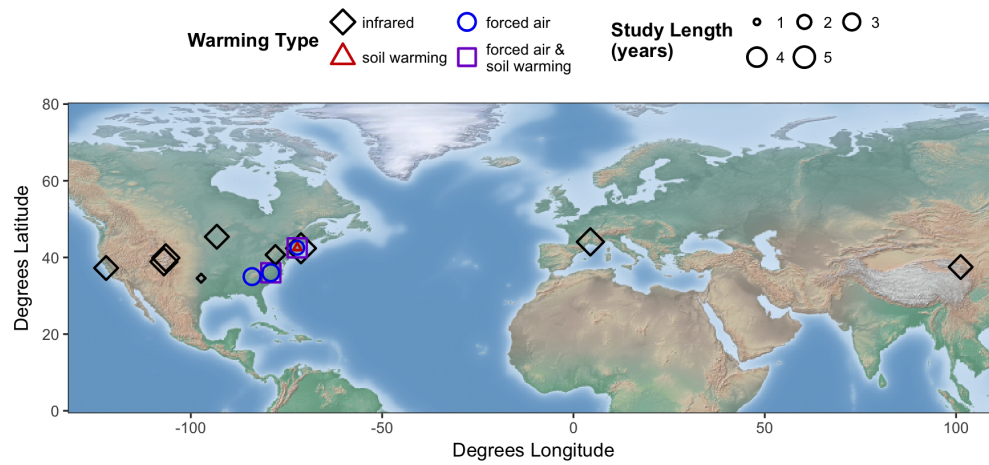


Figure S1: Climate data from 15 climate change experiments in North America, Europe, and China are included in the MC3E database and analyzed here. See Table S1 for details.

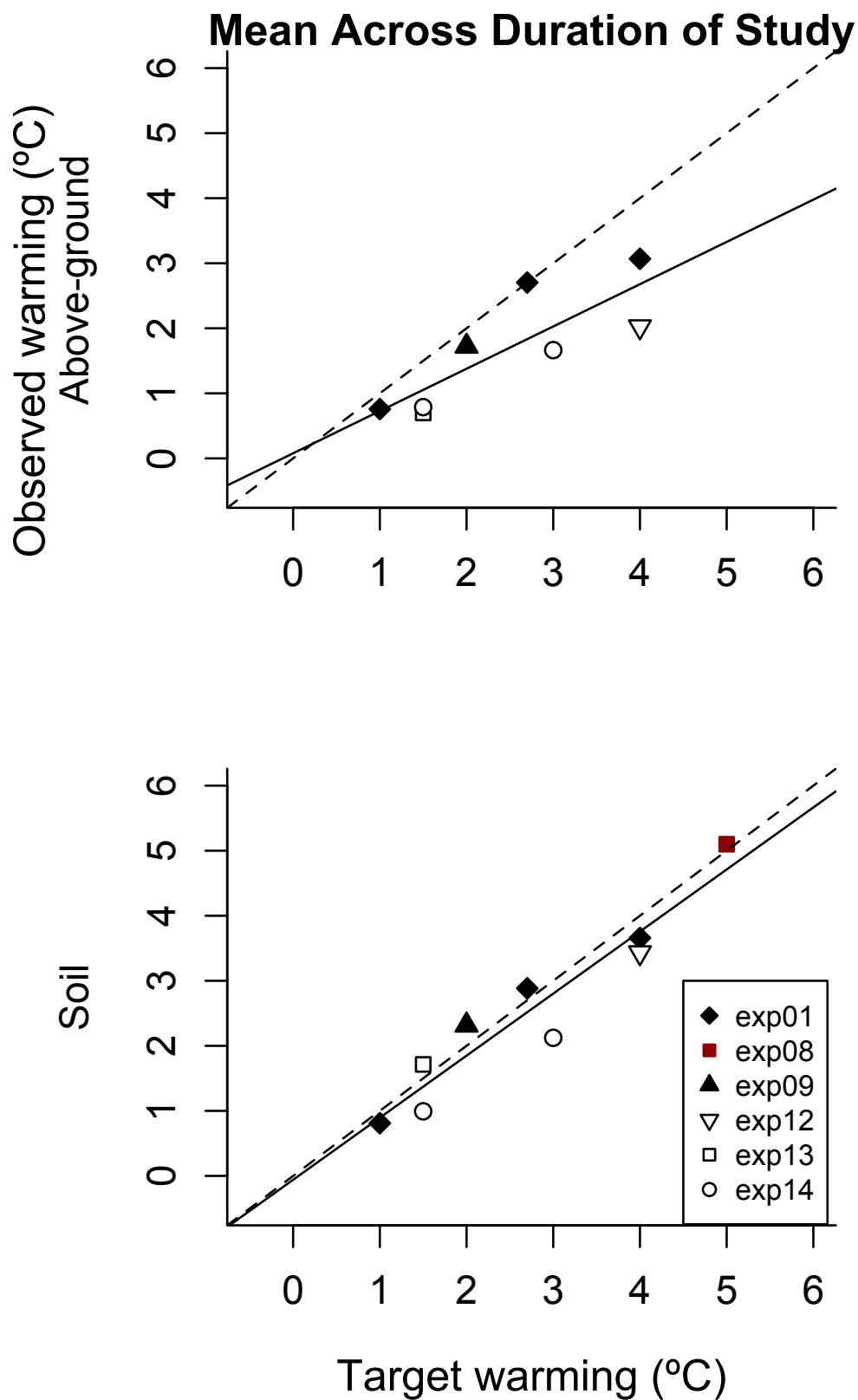


Figure S2: **Mean observed warming, for above-ground and soil temperatures**, excluding data from plots that manipulated precipitation. Above-ground temperature includes air, canopy, and surface temperature. Points represent the difference between treatment and control plots averaged across all plots 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.

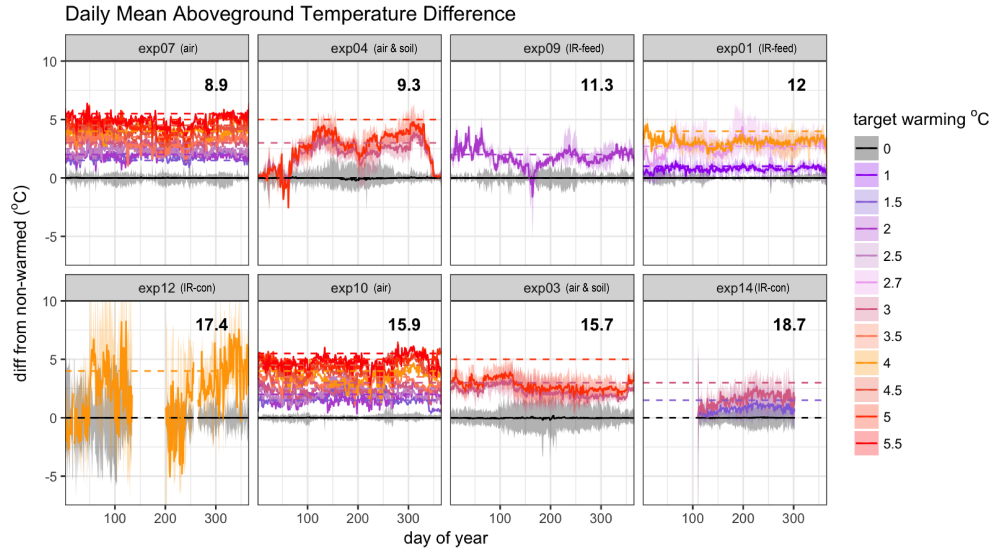


Figure S3: **Deviations in daily observed warming from mean control above-ground temperature for 8 study sites**, excluding data from plots that manipulated precipitation. We show above-ground temperature, which includes studies that measured surface, canopy, and air temperature (see Table S1). Solid 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.