



1300 Centre Street
Boston, MA, 20131

July 13, 2018

Dear Dr. Chase, and members of the editorial board:

Please consider our paper, entitled “How do climate change experiments alter plot-scale climate?” for publication as a Review & Synthesis in *Ecology Letters*. This manuscript is a revised version of manuscript ELE-00455-2018. We have incorporated the suggestions of the referees and editor by revising portions of the main text, figure and table legends, tables, and figures in the manuscript. We include a point-by-point response to the reviewer and editor comments.

As you may remember, our paper addresses a major need in ecology: improved understanding of the ways that active-warming field climate change experiments alter microclimate. We address this through creating and analyzing a new database of daily air and soil temperature and soil moisture data. We find that experimental climate change results may be interpreted in misleading ways, especially through the common practice of summarizing the mean changes in temperature across treatments. We show that such methods mask variation in treatment effects over space and time. We also find that indirect treatment effects, which are rarely interpreted with biological responses, may lead to under- or over-estimation of biological responses. We describe a case study of plant phenology, in which a mean-focused analysis, ignoring non-temperature effects, leads to inaccurate quantification of species’ sensitivities to changes in temperature. We present recommendations for future experimental design, analytical approaches, and data sharing to improve climate change experiments.

In the latest round of comments, both reviewers suggested changing our title, which we have done. Beyond this, Reviewer 2 was satisfied with the manuscript, but Reviewer 1 had several additional suggestions. This reviewer suggested that we needed to more carefully separate the different warming techniques used, and that our language still seemed to have a bias regarding the indirect effects of warming. Reviewer 1 also suggested we add additional details of comparisons made throughout the paper, and that we needed to more clearly express the sample size in our meta-analysis, which the reviewer worries is too low to make general claims about particular warming types. Please see below for details on how we have worked to address these comments.

We thank you for considering this revised manuscript for *Ecology Letters*.

Sincerely,

Ailene Ettinger Postdoctoral Fellow, Arnold Arboretum of Harvard University & Biology Department,
Tufts University

Reviewer Comments are in italics. Author responses are in plain text.

REVIEWER 1

The authors have worked hard to improve the manuscript and they have succeeded in many ways. I applaud their hard work and accomplishment. To me the most valuable part of the paper is the idea that realized warming, rather than an average treatment value, should be explored and reported, as well as, or perhaps instead of the mean warming. However, I still think

(1) the authors paint all different techniques (treatment types) with the same brush far too often (either intentionally or by failing to let reader know sufficiently which treatment types are involved in a given point) and far too centrally in the paper,

We appreciate the reviewer's concern about how the various designs may impact our findings. We have addressed this in two ways. (A) Improving the description of our statistical methods, noting that they give an overall estimate while controlling for other site-level differences (as possible) through our use of mixed-effects models. Thus, though we agree that our results are clearly dependent on the data—and thus warming designs—available for analysis, our methods our designed to give an overall best-estimate across designs. We have clarified this point (Lines 89-92). (B) We have added additional details about different treatment types throughout the figure and table legends and supplement as follows:

1. For all analyses, we now summarize the number of studies of each warming type in the description of the analysis performed (see Supplemental Materials, e.g., in the “Analyses of daily temperature range” section, we write “These studies included infrared (n=4), forced air (n=2), and combined forced air and soil cable (n=2) warming types.”).
2. For figures with points plotted (i.e., Figures 2, 3, S2), we use different colors to plot points for different warming types, in a consistent way throughout the manuscript (infrared=black, forced air=blue, soil warming=red, combined forced air and soil warming = purple). To improve clarity, we added to Figure legends when particular warming types are missing, e.g., in Figure 2, the legend now says “Black symbols represent studies using infrared; red represents soil warming cables (only exp08); no studies with forced air heating used a blocked design.”
3. For figures with lines showing different warming treatments (i.e., Figures 1, 4), we list the warming type for each experiment on its associated panel (infrared=“IR”, forced air=“air”, soil warming=“soil” combined forced air and soil warming =“air&soil”).

(2) also still carry what I see as ‘bias’ regarding the indirect effects of warming on water at spots. These come through with specific examples (noted below) but also in the overall language used. Although not as inaccurate nor insulting as in the original version, there is still more work that can be done here.

We thank the reviewer for sharing this perspective, and have made additional changes in our language (see below for details, responses to ‘some additional points’ 2-4).

(3) There are also quite a number of places where exact details of comparison are not made clear to reader, or at least not easily.

We have added additional details of comparisons.

(4) Given low numbers of experiments, I am not convinced that including experimental technique as a term in the analyses can do a good job of discerning the effect. Thus I would be circumspect about how such differences are reported. With n=1 or 2 for all experimental approaches other than infrared, I think the statistics say more about individual experiments than treatment approaches overall. And for IR, unless I miscounted, of the 9, six did not control temperature and three did (via feedbacks); shouldn't those be viewed separately, as by default the constant wattage experiments can not but help (at least from first principles) be more variable in temperature elevation than an experiment that tries to control

that elevation. In essence, although I compliment the author attempt to make a quantitative analysis of warming experiments, low replication still worries me greatly, and I think the language used in the paper should reflect that this is not a strong test (which is not the authors fault) of overall warming experiments or any given technique. Remember, there are five approaches: infrared with feedback control (n=2 or 3 as one is listed as both, infrared with constant wattage (n=6 or 7), forced air with feedback control (n=2), forced air + soil warming (n=2); and soil warming (n=1).

We thank the reviewer for these detailed, thoughtful comments. We agree that a higher sample size would be ideal, and wish that more data were available. We chose not to separate out constant wattage from feedback control in our analyses because of the low sample size. As the reviewer has correctly pointed out we are limited by low sample size and we considered this in how finely we divided the designs. We used a lower threshold of three observation units, in determining how finely to divide groups: if there were less than three different studies in a group, we either merged it with a similar group (e.g., forced air studies were merged with forced air plus soil cable studies, Table S16), or we did not use the grouping level in analyses. For this reason, most of our analyses do not include warming type.

We appreciate the reviewer's concerns that differences could be due to other differences between the studies included; however, we do account for this in our model structure, and we now more clearly highlight this (e.g., Lines 89-92). These and other methodological differences among experiments are accounted for in our models by including a random effects of 'site' in all models; the models the reviewer is concerned about thus partitioned variance due to method versus site.

In going through these revisions of this manuscript, we realized we needed additional clarity about the methodology used for the study that was listed as both feedback and constant wattage (exp01). The author, who is a coauthor on this paper, described how feedback was used to alter warming control in different treatments, and clarified that all warming control in their study should be considered 'feedback;' we have updated Table S1 to reflect this.

Regarding point 2 above, I remain somewhat unhappy with the approach taken by the authors vis--vis moisture impacts. In fact, I feel they have partially taken to heart strong suggestions of two of the three original referees, but retain what seems to me to be a strong bias in favor of unrealistic experiments (and likely in the view of original referee 1, unsophisticated) all with the goal of being able to isolate temperature effects from indirect moisture effects. See specific comments below.

Please see our responses to the specific comments below (responses to points 2-4 under "some additional points").

Not sure I am wild about the term "secondary". Why not use term "indirect" as it seems more accurate? Or if the authors don't like that, how about the term "non-temperature" effect. It is clunky, but most accurate. This term should also be defined at its first usage, even if in Abstract, as any effect other than on temperature.

We thank the reviewer for expressing concern over the term "secondary" and for suggesting alternative terms. We have removed "secondary" throughout our manuscript. We now use either "non-temperature" (in the context of experiments or analyses focused only on warming treatments) or "indirect" (when discussing treatments that include temperature and precipitation manipulations).

Authors need to do a better job of making clear to readers which techniques (warming types) are used in each figure and pointing this out to readers. Unless I read incorrectly, figure 2 includes no data for forced hot air experiments, and figure 3 includes ONLY data or such experiments (plus a soil warming experiment). Thus these figures are almost entirely about different experiments and different techniques. This needs to be made much much much clearer to the readers.

We thank the reviewer for pointing out the need for more clarity about this! The reviewer is correct

that Figure 3 includes data ONLY from forced hot air experiments, and Figure 2 includes no data for forced hot air experiments. Unfortunately, none of the forced hot air experiments included in the MC3E database used a blocked design (required to be included in Figure 2), and no infrared experiments for which we had data included both structural and ambient controls (required for Figure 3). This is a frustrating aspect of the available data; we would like to be able to compare across all warming types. We have worked to clarify the different warming types used in each analysis in the following ways:

1. For all analyses, we now summarize the number of studies of each warming type in the description of the analysis performed (see Supplemental Materials, e.g., in the “Analyses of daily temperature range” section, we write “These studies included infrared (n=4), forced air (n=2), and combined forced air and soil cable (n=2) warming types.”).
2. For figures with points plotted (i.e., Figures 2, 3, S2), we use different colors to plot points for different warming types, in a consistent way throughout the manuscript (infrared=black, forced air=blue, soil warming=red, combined forced air and soil warming = purple). To improve clarity, we added to Figure legends when particular warming types are missing, e.g., in Figure 2, the legend now says “Black symbols represent studies using infrared; red represents soil warming cables (only exp08); no studies with forced air heating used a blocked design.”
3. For figures with lines showing different warming treats (i.e., Figures 1,4), we list the warming type for each experiment on its associated panel (infrared=“IR”, forced air=“air”, soil warming=“soil” combined forced air and soil warming =“air&soil”).

Some additional points, in no particular order.

1. I think the title will be baffling to many readers. Even having read the initial paper, when I saw the new title my brain immediately asked how warming experiments can influence “local climate”, which to me is the climate of the overall site, not within the experimental treatment.

We have changed the title to “How do climate change experiments alter plot-scale climate?” which we believe addresses the concerns of the reviewer. While we appreciate the suggestion of “microclimate” we feel it is not a well-known term among many ecologists (many we asked assumed it meant the climate of a much smaller scale than we study here) and thus believe “plot-scale” is an accurate representation of our scale of study, while also being recognizable to most readers.

2. How could soil cable structural control increase air temperature (Fig 3)??? Or decrease soil moisture?? What is the mechanism?

The reviewer brings up some interesting mechanistic questions. We did not observe an effect of soil cable structural controls on air temperature; the one study (exp08) that used soil cables alone did not measure air temperature (only soil moisture is shown for this study in Figure 3, because minimum and maximum soil temperature were not available; only mean soil temperature was available). Figure 3 identifies the temperature type (above-ground versus soil) and heating type. As for the decreased soil moisture, it is possible that soil disturbance during the installation altered water flow to the area, which in turn may alter soil temperature and moisture. We discuss possible mechanisms in Lines 184-186, where we write “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.”

3. Lines 225-228. “Soil moisture can be difficult to measure, with high spatial and temporal variation (Famiglietti et al., 1999; Teuling & Troch, 2005), but these results suggest that soil moisture is unavoidably affected by changing temperatures, even when active-warming experiments may not be explicitly designed to manipulate soil moisture.” I would argue that by default active warming experiments, unless otherwise specified, do explicitly and simultaneously include direct effects on temperature and indirect effects on soil moisture. Saying otherwise is like saying an elevated CO2 experiment is not

explicitly designed to manipulate soil moisture through reduced stomatal conductance due to elevated CO₂. Changes, such as in rising CO₂ or elevated temperatures, have direct and indirect effects in the real world. Whether warming experiments do a good job of representing such indirect effects is a different, but important questions, Whether warming experiments actually dry soils as would be the case in a warmer world is an excellent question and I wish the authors had addressed that.

We appreciate the reviewer's point that an implicit intention of active-warming experiments may include the indirect effects of warming, such as soil drying. We have reworded this sentence, which now reads "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 soil moisture is unavoidably affected by changing temperatures in active-warming experiments."

4. They go on to add "Warming and precipitation treatments, and their secondary 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). These community shifts may affect resource levels, such as moisture, carbon, and nutrient levels in the soil (McDaniel et al., 2014b,a; Harte et al., 2015) and feedback to affect microclimate (Harte et al., 2015). The presence of these feedback effects is both a strength and a challenge of climate change experiments. They may represent important and ecologically realistic effects that might not have been apparent without the in situ field experiment. Alternatively, they may represent artifacts that are unlikely to occur outside of an experimental context. Quantifying, interpreting, and reporting these non-temperature effects in experiments is critical to distinguishing this and to understanding mechanisms underlying observed biological responses to climate change." This makes sense although it would be nice to know what kinds of effects are likely realistic and which ones artefacts, and again, to know whether certain kinds of warming treatments are more likely than others to induce this.

We thank the reviewer for this point, and agree that distinguishing between realistic effects and experimental artifacts is a critical issue. We discuss this in two paragraphs in the "Ecological Implications" section (Lines 297-326). For example, we write in Lines 301-309:

For many important climatic and ecological metrics, experimental findings of abiotic and biotic effects appear to be consistent with observations. Altered above-ground daily temperature range (i.e., temperature minima changing more than maxima, Table S3) with experimental warming is consistent with observed changes in many places, at least for some time periods. Global minimum temperatures increased more rapidly than maximum temperatures from 1950-1980, reducing above-ground daily temperature range (Thorne et al., 2016; Vose et al., 2005). In addition, the acclimation response of leaf respiration to temperature (Aspinwall et al., 2016; Reich et al., 2016), responses of soil respiration to warming (Carey et al., 2016), and declines in soil carbon at one site (Harte et al., 2015), also appear to be consistent across experiments and observations.

We also suggest some aspects that may be artifacts. For example, we write (Lines 311-319):

In contrast, some responses documented in climate change experiments may not be in line with future climate change, and thus need explicit analyses and cautious interpretation. For example, soil drying in conjunction with future warming is forecasted in some regions, such as the southwestern United States, mainly because of reductions in precipitation and increased evaporative demand with warmer air (Dai, 2013; Seager et al., 2013). The north-eastern United States, on the other hand, has been trending wetter over time (Shuman & Burrell, 2017), even though temperatures have warmed. Future changes in soil moisture are uncertain, and likely to vary by region, season, and even soil depth (Seager et al., 2014; Berg et al., 2017). Thus, it is not safe to assume that the soil drying observed in warming

experiments is necessarily likely to occur with future warming.

The reviewer brings up a question that would be incredibly useful to understand: whether certain kinds of warming treatments are more likely than others to induce artifacts. We wish we were able to address this; however, given the small sample size of non-infrared studies in our database (e.g., 3 forced air studies, 1 soil cable study, and 2 combined forced air and soil cable studies), we are not able to make strong statements about what kinds of warming treatments are more likely to induce particular artifacts than others. We do point out differences in measured variables by warming type, when possible. For example, we report the amount of drying by warming type (Lines 228-231): “We found that experimental warming reduced soil moisture across all warming types, with substantial variation among experiments (Figure 5, Table S15). The drying effect varied by warming type (-0.80% for infrared versus -0.33% for forced air, per degree of target warming, Table S16).” We also statistically account for differences among studies by including a random effect of study (‘site’ effect) in all our analyses.

5. *The authors wrote “We have highlighted a suite of factors that complicate interpretation of climate change experiments. These largely non-target alterations, analogous to the “hidden treatments” described by Huston (1997) in biodiversity experiments, are likely to have biological implications for many of the responses studied in warming experiments (e.g., Figure 5).” But what if researchers want to know what the direct and indirect effects of warming are on target organisms? Isn’t that the goal of many (who knows how many; half? Two-thirds? Three-fourths)) warming experiments? Of course, in such cases research would and do attempt to disentangle the relative effects, but treating 5a as the desired ‘target’ of the research seems wrong, or at the very least a subjective choice. Use of the word ‘non-target’ is (as with first submission) insulting and inaccurate.*

In light of the reviewer’s concerns, we have reworded the sentence, to avoid using the term “non-target.” The sentence now reads: “These indirect effects are similar to the “hidden treatments” described by Huston (1997) in biodiversity experiments, and are likely to have biological implications for many of the responses studied in warming experiments (e.g., Figure 5).”

6. *Authors wrote: “Decreasing soil moisture has a delaying effect on budburst phenology, opposing the advancing effect of rising temperatures (Figure 5b); thus the effect of temperature is underestimated when moisture is not included in the model. This example shows how the common method of using target warming alone to understand biological responses is likely to yield inaccurate estimates of temperature sensitivity in warming experiments. In this case, the underestimation may be substantial enough to account for the previously observed discrepancy between observational and experimental phenological responses to warming (Wolkovich et al., 2012), though further investigation is required.” Did the original papers in questions ignore the indirect effects of warming on soil moisture in drawing conclusions? If so, then this is a very valid point and this omission should be pointed out. If original authors did mention such effects, it is ok to revisit them, but please make clear that original authors already were paying attention to this point. This is not clear as currently written.*

We thank the reviewer for highlighting this lack of clarity. Of the five experiments included in our budburst analysis, four of them have resulted in published papers on warming effects on phenology (exp01, exp03, exp04, exp10; Clark et al 2014a, Clark et al 2014b; Marchin et al 2015; Polgar et al. 2014). None of these previous analyses included soil moisture as an explanatory variable for phenology; warming treatments and/or temperature variables (including growing degree days) were included as explanatory variables in their phenology models and discussed in drawing conclusions. Soil moisture has been explicitly analyzed in some non-phenology publications from these experiments, however (e.g., tree sap flow in Juice et al 2016, soil and rhizosphere respiration in Suseela and Dukes. 2013).

We have rewritten this section to improve clarity and to add an additional reference (a study that compared observational to experimental sensitivity to warming in plant phenology; Polgar et al 2014).

It now says (Lines 290-297):

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 is likely to yield inaccurate estimates of temperature sensitivity in warming experiments. In this case, the underestimation may be substantial enough to account for previously described discrepancies between phenological responses to warming in observational versus experimental studies (Wolkovich et al., 2012; Polgar et al., 2014), though further investigation is required.

7. Authors wrote: *“In contrast, some responses documented in climate change experiments may not be in line with future climate change, and thus need explicit analyses and cautious interpretation. For example, soil drying in conjunction with future warming is forecasted in some regions, such as the southwestern United States, mainly because of reductions in precipitation and increased evaporative demand with warmer air (Dai, 2013; Seager et al., 2013). The northeastern United States, on the other hand, has been trending wetter over time (Shuman & Burrell, 2017), even though temperatures have warmed. Future changes in soil moisture are uncertain, and likely to vary by region, season, and even soil depth (Seager et al., 2014; Berg et al., 2017). Thus, it is not safe to assume that the soil drying observed in warming experiments is necessarily likely to occur with future warming.”* This is a bit of a red herring as worded. It is the overall experiment that may not match future moisture conditions, not just the warmed treatments. This can be incorporated into a design but with a cost (either huge costs of replicating factorial combinations of current and future conditions for temperature, rainfall and CO₂; or cost of being unable to disentangle mechanism if a treatment like ‘iii’ below is used). Higher ET will always happen in a warmer context with all else being equal. If (and this is uncertain) experimental warming increases water loss in a realistic way, then given the rainfall in any given experiment, the warming treatment indirect effect on soil moisture is ‘spot on’. This follows from basic physics and will happen in any future warmer condition. It is the ambient rainfall (and perhaps total atmospheric water content) condition that may not reflect future climate (which may have less or more precipitation or water content). I think our ability to predict future rainfall and/or future total water vapor (which drives VPD when combined with temperature) is so poor that no one seriously aims a rainfall manipulation to mimic future rainfall or tries to realistically manipulate total water vapor, but instead to provide a scenario of one plausible future. Instead warming experiments superimposed on today’s rainfall in any given year in any given site do assess the direct and indirect effects of warming given the ambient rainfall experienced. If warming effects are contingent upon future VPD and/or precipitation (which is likely), the unwarmed plants would also respond differently to such changes. Are the authors arguing that if someone is confident that location Q is going to become 2C warmer and have 20% less precipitation one could design a (non-factorial) treatment that included both of those- so that today’s temperature were coupled with today’s rainfall in treatment A and tomorrow’s temperatures were couple with tomorrow’s rainfall in treatment B? We should throw ambient CO₂ into treatment and elevated CO₂ into treatment B. Such a contrast of treatments would be ecologically oriented and more realistic than just changing one factor and holding the others constant, but mechanistically extremely difficult to interpret. I worry that I was not entirely clear above, so here as an example are some design choices (hypothetical, some may be near impossible)

i. Increase temperature, hold soil moisture and VPD constant across treatments- this assesses effect of rising temperature independent of indirect effects of warming on soil moisture and VPD, given whatever any given years rainfall happens to be.

ii. Increase temperature, allow soil moisture and VPD to change as a function of altered ET - this assesses direct effect of rising temperature and indirect effects of warming on soil moisture and VPD; all given whatever any given years rainfall happens to be.

For both i and ii, if experimental years are wetter or drier than future years, neither measured response (i or ii) may be an accurate predictor of future change, as the observed warming effect of both ambient and warmed temperatures will probably be contingent on moisture availability. In other words, the problem is not with the indirect warming treatment effects, it is with the base climate hydrology (rainfall and total air water vapor) in both treatments.

iii. Increase temperature, allow soil moisture and VPD to change as a function of altered ET from increased temperature, and also manipulate rainfall and total atmospheric humidity to better match conditions in some future target year (assuming we have some idea what this might be). Conceptually this is like also adding elevated CO₂ to the warmed treatment and not to the ambient temperature treatment (i.e. without factorial treatments). This experiment does the best job of testing effects of future conditions on target responses (most ecologically realistic), but is the most difficult to cleanly interpret, as multiple factors are changing together.

We agree with the reviewer that replicating future soil moisture conditions, and interpreting the results of experiments that claim to do so, is an enormous challenge, made particularly difficult by our uncertainty in forecasting. We wish to frame this as an ongoing challenge for ecologists and climate scientists, and we know that many ecologists have been aware of this challenge for decades. Our main argument is not that ecologists need to alter the way that they manipulate soil moisture; rather, we wish to suggest that an important step is for soil moisture to be explicitly analyzed and incorporated into analyses of ecological responses in warming experiments. We have added the following sentence to Lines 319-322 to make this point clearer: “The uncertainty associated with forecasting changes to soil moisture makes replicating future water availability regimes in climate change experiments especially challenging; one way to meet this challenge and make predictions– even given high uncertainty– is to estimate soil moisture effects in climate change experiments.”

8. Lines 318 to 330 make sense, but I am not quite clear on how this relates to the topic and specifics of the analysis provided in the paper.

We understand the reviewer’s point and agree that this paragraph was too long, as previously written (now Lines 327-339). This paragraph was added to address concerns of one reviewer in the earlier version of this manuscript, and we feel that it addresses important issues regarding differences between experimental and observational data. In light of the reviewer’s comment, we have removed several phrases and a whole sentence from this paragraph to shorten it.

9. Table S1 should have column identifying whether chambers were used or not. Table S2 should summarize by warming type the performance metrics.

We thank the reviewer for pointing out that this information would be useful to the reader. For Table S1, we have added text in the Supplemental Materials’ “Analysis of effects of infrastructure on experimental microclimate” section: “Note that all studies that employ forced air warming utilize chambers, whereas the other warming types did not utilize chambers.” We also added the following to the legend of Table S1: “All studies that employ forced air warming utilize chambers, whereas the other warming types did not utilize chambers.” As this is already a very large table and chamber presence is perfectly associated with warming type (which is already in the table). We hope the reviewer will understand our choices here, which we believe balance important information with a readable table.

10. Not clear from Fig 4 legend whether the percent soil moisture deviation is the percent of the percent or the absolute difference in VWC; (i.e. if the two values were 0.4 and 0.3, the former would be 25%, the latter would be 0.1 (or 10% if VWC were expressed as a percent rather than a fraction).

We thank the reviewer for pointing out this lack of clarity. The absolute difference is shown. We added the following text to the legend, to clarify this: “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.”

12. *Here I have an issue with wording in abstract. Abstract line 7-8 Our synthesis showed measured mean warming in plots with the same target warming can vary by 3°C or more among blocks. Tell reader the mean of median difference not the maximum. And say which treatment type was used.*

We have changed the abstract, which now reads: “Our synthesis showed that measured mean warming, in plots with the same target warming, varied by 1.6°C, on average, among blocks within a study.”

13. *Figure 6. Legend does not state clearly whether the differences were between treatment and structural control, not treatment and ambient control.*

We thank the reviewer for pointing out this oversight. We now state in the legend: “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).” We have also added the following to the Supplemental Materials “Analysis of budburst phenology” section: “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).”

14. *Table S1 legend. Spell out for reader what ‘structural’ vs ‘ambient?’ control means.*

We thank the reviewer for pointing out the need for this information in the legend. We have added to the legend “...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)...”.

REVIEWER 2

I have always been a fan of this paper, and I was pleased to see how thoroughly the authors addressed the review comments. I think this will be an important addition to the literature and a must read for anyone setting out to establish a climate change experiment. My only suggestion is to change the title to indicate : “How do climate change experiments alter local microclimate?”, but I do not feel strongly about the change.

We thank the reviewer for these kind words, and the suggestion. We have changed the title to “How do climate change experiments alter plot-scale climate?” which we believe addresses the concerns of the reviewer. We discuss our reasoning for not using “microclimate” above (see R1. pt 1 under ‘Some additional points’).

EDITOR

Both reviewers commended the additional work and revisions made by the authors, which has resulted in significant improvements to the manuscript. Both identify some highly important and relevant messages in the paper regarding best practices in climate change experiments.

R1 identified remaining concerns that can be addressed through further clarification and justification within the text. In particular, how might the low number of replicates for each treatment type within the quantitative analysis affect the conclusions about the indirect effects of warming on soil moisture? The mixed effects model used in the study should be described in the text in order to better justify the approach used to generate these conclusions.

We thank the editor for this suggestion. We have added details on our mixed-effects models, and highlighted that they should account for some of the methodological (and other) differences among sites, to the main text in several places:

1. Lines 89-92: We now say “We examined how these experiments altered microclimate, using mixed-effects models that allowed for inherent differences among studies (through a random effect of study on the intercept), while also estimating various across-study effects, such as degree of warming or warming type.”

2. Infrastructure analyses (Lines 157-159: We now say “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..., accounting for methodological differences among studies by including a random effect of study (see Supplemental Materials for details).”
3. Soil moisture analyses (Lines 225-228): We now say “Of the 15 experiments in the MC3E database, we examined the 12 that continuously measured and reported soil moisture. We included target warming, warming type, and their interaction as predictors (excluding data from plots with precipitation treatments) and accounted for other differences among studies by including a random effect of study (see Supplemental Materials for details).”
4. Budburst analyses: Lines 266-272 now say: “We fit two separate mixed-effects models, that differed in their explanatory variables: one used target warming and one used measured climate. Both models had budburst day of year as the response variable, and both included random effects of study (which modeled other differences between studies, that may have affected phenology), year (nested within study, which modeled differences due to weather variability among years that may have altered phenology), and species (which often vary in their phenology, all with an intercept-only structure, see Supplemental Materials for details).”

R1 also suggests that the authors address the feasibility and costs of implementing an effective soil moisture experiment. Related to this point, R1 requested clarity over whether the original authors of the cited studies mentioned warming effects on soil or truly ignored them. A revised manuscript should recognise current limitations on our ability to predict and therefore replicate future water availability regimes in climate change experiments; the current revision still implies that the lack of information on soil water availability is an oversight instead of an on-going challenge.

We thank the editor for this suggestion, and particularly appreciate the framing of soil moisture as an “on-going challenge.” We have adjusted the topic sentence of the paragraph and added the following sentence to reflect this to Lines 319-322: “The uncertainty associated with forecasting changes to soil moisture makes replicating future water availability regimes in climate change experiments especially challenging; one way to meet this challenge and make predictions– even given high uncertainty– is to estimate soil moisture effects in climate change experiments.”

Finally, both reviewers suggest that the authors reconsider the use of 'local' climate in the title; R2 suggests the authors use 'micro' instead.

We have changed the title to “How do climate change experiments alter plot-scale climate?” We discuss our reasoning for not using “microclimate” above (see R1. pt 1 under ‘Some additional points’). If the editor feels strongly that an alternative title would be better we can change it further.