



1300 Centre Street
Boston, MA, 20131

October 17, 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.R1. We have incorporated the suggestions of the referees and editor by adding a new table, modifying figures, and revising the main text. 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 most recent round of comments, Reviewer 1 felt that our manuscript offered an important contribution to the literature and suggested no changes. Reviewer 2 felt that two main issues required revision: 1) variation among experimental warming techniques (specifically feedback versus constant wattage approaches) was not made clear enough; and 2) “indirect effects” were presented in a biased way. We have extensively modified the manuscript to address these concerns: we made changes to 4/6 figures, changed the abstract and conclusion, and made changes to all sections of the main text. We also have added a new box, a new table, and a new figure, all focused on warming techniques, to address these concerns, and we reference these new presentation items throughout the manuscript.

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

Sincerely,

Ailene Ettinger Postdoctoral Fellow

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

REVIEWER 1 (called Reviewer 2 in the previous version and in the editor's comments below).

This manuscript continues to be improved thanks to the astute comments from the reviewers and the editor. I maintain that this is very important contribution to the literature, and will be of great interest to the global change experiment community.

Thank you very much!

REVIEWER 2 (called Reviewer 1 in the previous version and in the editor's comments below).

Unhappily, I am somewhat disappointed with this revision. I was frustrated in part because the authors have paid a lot of attention to certain details and issues and largely ignored others, resulting in a text (starting with the abstract) that in my opinion would be a disservice to our field if published as is. Although some improvements have been made, many of my concerns from the prior review remain partially or largely intact. The authors made modest and often cosmetic changes to try to assuage criticisms, seemingly trying to change their message as minimally as they could to retain their "story" while appeasing referees.

We greatly appreciate the time and effort the reviewer has dedicated to improving this manuscript. We are sorry to hear of the reviewer's disappointment, especially because we did seriously consider the reviewer's valuable comments and worked to address them in the previous revision.

We realize now that perhaps we were not as clear and thorough as we could have been in describing the substantial new analyses, new data presentation, and other substantial time and effort aimed at addressing suggested changes in the previous round of edits. We have worked hard in this revision to fully address the reviewer's concerns, and to detail our efforts clearly and thoroughly in this response.

The authors did make some changes in terms of how they discuss moisture, but largely ignored the key issues involving different experimental techniques for aboveground warming—specifically ignoring the important differences between constant wattage and feedback control, and more generally talking about issues largely specific to only certain techniques (e.g., control structure effects) as if they applied broadly to all warming techniques and experiments. This failure to differentiate among warming technique to me is an enormous problem. As one example, giving the impression in the Abstract that the typical warming experiment averages 1.6C different degree or extent of warming from block to block paints all warming experiments with one brush when this problem is much greater for one outmoded approach (constant wattage). It is like reviewing treatments for cancer Q and reporting only that they average 25% serious side effects. This is useless information if one of two equally common treatments (but now recognized as vastly inferior, in part for this very reason) averages 42% and another 8%

We thank the reviewer for detailing the concerns about differences among different warming techniques, and pointing out the need to differentiate control type (constant versus feedback). We realize now that, in the previous review, we had misunderstood the reviewer's primary concern of important differences between constant wattage and feedback control. We had thought that the reviewer's primary concern was with respect to warming type (infrared, forced air, soil cables, and forced air plus soil cables). In the previous revision, we added warming technique to figures, conducted new analyses incorporating warming technique, and added discussion of the different warming techniques throughout the main text.

In this new version of the manuscript, we have additionally distinguished warming control (feedback versus constant wattage) to figures and created a new box that describes the importance of variation in different warming methodologies, and a new table that summarizes differences in observed warming by warming technique AND warming control. We detail this further below.

We agree with the reviewer that the fundamental physics of feedback versus control wattage would be expected to lead to lower block to block variation with feedback control. However we do not find

this relationship in our data. We explain this in detail below (see responses to issue 1, a-f). We have adjusted the abstract to address the importance of understanding variation due to warming methods (Lines 8-11), where we write “Our synthesis showed that measured mean warming, in plots with the same target warming, differed by 1.6°C, on average, with high variation— maximum differences ranged from 0.03 - 4.87°C across studies, which include variation due to warming design and myriad ecosystem attributes.”

Here I discuss the two main issues (adequately handling “techniques” and the topic of indirect effects).

1. *Issue of variation in experimental aboveground warming techniques*

- (a) *I believe it will be easy for readers to come to erroneous (in the sense of over-generalized) conclusions because of the way in which this paper is presented. It is not the authors fault that there are low numbers of experiments with each of the three main types of treatments (infrared constant wattage, infrared feedback, chambered [forced air]). The authors admit the data are sparse and problematic in terms of coming to conclusions; and because there are few replicates of each, there is insufficient data to statistically test whether and how technique matters to the statistical conclusions. They use that as an excuse (a valid one) for not differentiating such differences statistically. However, they persist in talking about and making conclusions about warming experiments as if they all have similar problems, rather than specifying which ones have the particular problems they highlight, and stating that differences in techniques mean one can not, and should not, even try to come to generic conclusions about all warming experiments.*

Giving the impression that there are common problems to most warming experiments is (in my opinion) more harmful than helpful to readers, because different kinds of warming treatment technologies have vastly different characteristics and advantages and flaws. The techniques are different enough to know without data what those might be, and when data are available, data differences are striking enough, that this needs to be done for the paper to be a net positive rather than net negative contribution to our field. I understand that the authors did not set out to review the advantages and disadvantages of different warming techniques, but given the huge apple-to-orange differences, the only useful way to present information is by being crystal clear which technique is being discussed and/or included in which kind of information.

We thank the reviewer for detailing the concerns about differential effects of different warming techniques, and pointing out the need to differentiate control type (constant versus feedback). In the previous review, we sought to address a similar concern by adding warming type (infrared, forced air, soil cables, and forced air plus soil cables) to figures and discussing in detail effects of the different warming types. In this new version of the manuscript, we have additionally distinguished warming control (feedback versus constant wattage) in figures: we have now changed Figures 1, 3 & 4 to address this important point. We have added a new figure (Figure 2), which shows photographs of different warming types. We also discuss the critical differences among warming technologies in the following places of the main text:

- i. Lines 168-177
- ii. Lines 188-194
- iii. Lines 199-206
- iv. Lines 236-238

We have also added a new table (Table 1) and a new box (Box 1) in the main text that summarizes the five different combinations of warming techniques and control types found in our database. These include infrared feedback, infrared constant wattage, forced air feedback, soil cable feedback, and forced air plus soil cable feedback. The table summarizes the

quantitative differences in warming among these types in the MC3E database (despite the low sample sizes). It also summarizes other non-temperature effects of warming treatments documented in the present manuscript and in other published literature. The box discusses the importance of additional research on warming methodologies. We now reference this Box eight times in the text and we discuss the need for additional research in the abstract and on 113-115, Lines 145-148, Lines 217-219, Lines 365-368, Lines 673-675, and Lines 694-696.

We also agree with the reviewer that we should be clear about which methods are included in which analyses and have worked hard to make sure this is clear across the paper. We now clearly state in the text, figures, and supplement which methods are included in which analyses. In addition, we agree that we our previous version may not have been as clear about the *implications* of these different methods throughout the manuscript. We have added text highlighting the particular warming methodologies and how this may affect our findings, when relevant. For example, at lines 186-88, where we and previously written “Infrastructure materials may shade the plots, reduce air flow, reduce albedo relative to surroundings, or otherwise change the energy balance”, we have added “, particularly in warming methodologies that employ chambers (i.e., 4 of the 5 studies included in the above analysis).”

For example, in the Abstract, their first main take-home point about their results is the following “We find that the common practices of analyzing primarily mean changes among treatments and analyzing treatments as categorical variables (e.g., warmed versus unwarmed) masks important variation in treatment effects over space and time. Our synthesis showed that measured mean warming, in plots with the same target warming, varied by 1.6°C, on average, among blocks within a study.” This comes across as a blanket statement about warming experiments, but fails to point out that this is an enormous problem for constant wattage experiments (e.g. Experiments 12, 13, and 14)[as has been discussed previously in the literature], and a real but much smaller one for feedback control experiments 1 and 9. This failure to differentiate between these techniques was both point 1 and 4 of my prior review, and was mostly ignored. A blanket statement like this could take on a life of its own and become the ecological equivalent of “common knowledge.” The point the authors make here in the Abstract is a good one- spatial variation matters, but giving the impression that all warming experiments have block to block variation of something on the order of 1.6C for ostensibly the same warming treatment does a disservice to the field, rather than helping. Eyeballing it from Figure 2 upper left panel, it looks like the mean variation among each pair-wise block is roughly 0.4 to 0.7C for feedback experiments and although it is hard to see given so many overlapping points for Expt 14, if the mean reported was 1.6°C across all studies, the mean for just constant wattage is likely between 2.5 and 3.5C. Those are large, critically important differences among the two approaches and are entirely consistent with the physics and engineering of each approach. Constant wattage experiments must and will increase temperature less when vegetation is dense and especially when ET is high. So a block with denser vegetation and higher ET (for whatever reason) will have lower temperature elevation. This is problematic for constant wattage experiments with multiple treatments or sites, where warming increases biomass and ET more in some warming treatments than others. The authors of this EL submission could help readers by identifying such studies and while some authors of constant wattage experiments are very open and upfront about such issues, others ignore this issue entirely, so helping the reading community be aware of this is useful. In contrast, by design, feedback controls alter the wattage to try to maintain a constant temperature elevation. So it is not a surprise that performances of the two types of technique differ dramatically even within the limited data this EL submission has in hand. I think the authors MUST specify the variation among blocks for each experiment type (constant wattage; feedback) in the Abstract and in the main text and refrain from making blanket

statements about all warming experiments that ignore large differences among them. Additionally, even though the authors have focused on these 15 studies for which they had data, I think the authors need to at least look at other feedback or constant wattage experiments and, if block to block data are available, which they may not be, report on these in the Discussion to get a better feel for just how much variation on average such experiments include.

We thank the reviewer for suggesting ways to further help readers understand the important differences between sites and methodologies, such as constant wattage (exp05, exp06, exp11-14) versus feedback control for warming (all other studies, including exp01, exp02, and exp09 for infrared). To address the reviewer's concern about specifying variation by block in constant versus feedback warming in the Abstract, we conducted new calculations of block to block variation (Figure 3), by warming control type.

We expected, as the reviewer did, to find substantially greater variation among blocks in studies using constant wattage as compared to feedback control studies. However, we did not find support for this prediction. The constant wattage control studies have greater average variation (2.2 °C variation on average for constant versus 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, varied by as much as 4.87 °C on average; least: plots in exp13 with 1.5 °C of target warming, varied by 0.03 °C per °C of target warming). These results of course are not conclusive, especially because our sample size is quite low ($n=3$ studies for constant and $n=2$ studies for feedback studies with blocked designs). We discuss these new results in Box 1.

We wanted to address the reviewer's important point that there are differences between sites (and perhaps methodologies, though we do not have evidence to support this) while making sure that all statements in our abstract (and throughout the manuscript) are supported by data in the MC3E database. Therefore, we now say "Our synthesis showed that measured mean warming, in plots with the same target warming, differed by 1.6 °C, on average, with great variation: maximum differences ranged from 0.03 - 4.87." (Lines 8-10). Because the means between the two warming control types did not differ statistically, we did not feel that we could accurately state that there were differences by control type, although there was wide variation across (and within, in some cases) studies.

We addressed the Reviewer's suggestions in three additional ways: 1) We now distinguish control type, as well as warming technique, in Figures 1-5. 2) We have also added a new table (Table 1) and a new box (Box 1) in the main text that summarizes the five different combinations of warming techniques and warming control types found in our database (these include infrared feedback, infrared constant wattage, forced air feedback, soil cable feedback, and forced air plus soil cable feedback). The table summarizes quantitative differences calculated from the MC3E database (to save the reviewer and future readers from the need to "eyeball" points in Figure 3 (formerly Figure 2)) and described in the manuscript. The table also summarizes other non-temperature effects of warming that we found in published studies. (We welcome the reviewer to provide additional citations if we have missed important ones that should be added.) 3) We have added language throughout the manuscript to remind the reader that many effects may vary by warming technique, even if we do not have data available to quantitatively test this. For example, we added the following sentence to Lines 113-115: "When possible, we compare and contrast these factors across different study methodologies, such as infrared warming versus forced air chambers and constant energy output versus feedback control, because effects on microclimate may vary across these different methodologies (Box 1)."

- (b) *The same issue is relevant at lines 112-122 and lines 123-133. Again, although all treatment techniques may be unable to maintain target warming during periods of highest LAI and ET, constant wattage experiments by default will have more of a problem with this.*

We appreciate the reviewer's reminder that constant wattage experiments are likely to be particularly problematic at maintaining target warming over time, regardless of weather and other environmental conditions. To address this point, we have added the following sentence to Lines 134-136: "Differences between target and actual warming are likely to be particularly great for studies employing constant wattage, rather than feedback control (Box 1, Kimball et al 2008)."

- (c) *Lines 134-142 deal with the issue I have a problem with in the Abstract and here too it should be clarified how these two techniques will differ.*

We again thank the reviewer for pointing out the need for distinguishing between two critically different techniques for warming. We have added the following sentence to address the reviewer's concerns (Lines 145-148) and call attention to our new Box 1: "This variation in warming is substantial, as it is equivalent to the target warming treatment for many studies, and appears to vary substantially among sites, which differ in warming methodologies and environmental characteristics, though sample sizes for each distinct combination of methods are low (Box 1)."

- (d) *Lines 150- 204 (157-167 in particular)*

Here the effects of experimental infrastructure on conditions is highlighted, based on analyses of forced hot air treatment technology only. Although this detail is mentioned, it is buried in the text; thus the heading should highlight this: 'Chamber (forced hot air) heating alters microclimate.' This would help the readers keep in mind that each section of this paper highlights issues with DIFFERENT KINDS of heating systems. As written, each section highlights problems, and your average reader may come away from reading this paper thinking, "wow, all warming systems suffer from so many problems!"

We thank the reviewer for pointing out the need for additional clarity in this section, and other sections, specifically regarding the warming technology included in the analyses. The analyses in the section highlighted by the reviewer (now Lines 160-219, Lines 166-177 in particular) actually include both forced air (chamber) and soil cable warming only (non-chamber) types.

Unfortunately, no infrared studies included both ambient and structural controls, so we were unable to include them in the analysis. We do discuss in this section a separate analysis of infrared studies, which is suggestive of microclimate effects from infrastructure, likely via shading, which has been documented in other studies (see Table 1). Thus, because we discuss three distinct warming types (forced air, soil cable, and infrared) it would be inaccurate to have a heading of 'Chamber (forced hot air) heating alters microclimate.' To address the reviewer's concerns, we considered changing the heading to 'Forced hot air and soil cable infrastructure alters microclimate (infrared infrastructure untested).' This heading, however, seemed long and cumbersome, and infers that other designs do not have these issues, when instead we do not have data to assess these issues in other designs, which we state on Lines 173-177.

- (e) *Line 157-167 point out issues with structural controls. I believe those were all chambers? If so, they should have larger effects than open-air techniques. Note: SPRUCE has enormous chamber effects which the researchers do acknowledge.*

We thank the reviewer for pointing out that chambered structural controls may have dif-

ferent effects than open-air techniques. We now mention this point in Lines 194-197, where we say “These effects may be most dramatic in studies that employ chambers, rather than open-air designs, such as infrared heating (Aronson and McNulty, 2009).” Note that these are not all chamber studies: the studies included in the analysis in this section include four chamber studies (forced air, and forced air with soil warming) and one open-air study (soil warming only), as mentioned lines 169-172.

- (f) *I suggest that at the end of the paper the authors highlight which kinds of systems (forced-hot air, infrared constant wattage, infrared feedback control, etc) are likely to suffer from which of the issues they highlighted in the paper. Although their approach to this topic is laudable, to try to see what the available data teach us, in doing so they are comparing apples and oranges, and given modest sample sizes, the authors should use qualitative as well as quantitative comparisons.*

I also recommend that each subsection of the results name the kinds of warming treatments that are examined in that section, and that the authors constantly remind readers of which results apply to which apply to which treatment types.

We thank the reviewer for these suggestions. We felt that adding the names of all the different warming treatments in each section led to section headers that were a bit clunky (e.g., “Effects on microclimate vary over time and space for infrared, forced-air, and combined forced-air and soil warming”). Instead, we remind the reader throughout the results which treatments are included in which analyses, and we discuss any differences (or lack of difference) we observed across warming types within the text of each section. This includes the following places, among others:

- i. Lines 123-125 in the “Effects on microclimate vary over time and space” section, where we write: “These effects varied by site (Table S3), but we found no clear patterns by warming type (e.g., infrared versus forced air) or warming control (feedback versus constant).”
- ii. Lines 132-136 in the “Effects on microclimate vary over time and space” section, where we write “Both infrared heaters and soil cables fail to achieve target temperature increases during rainstorms... 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, Figure 2).”
- iii. Lines 142-148 in the “Effects on microclimate vary over time and space” section, where we write “These studies include two infrared with feedback control, three infrared with constant energy input, and one soil warming cable with feedback control experiments. We found that the amount of observed warming frequently varied by more than 1C (mean= 1.6C, maximum = 4.9C) among blocks (Figure 2, Table S6). This variation in warming is substantial, as it is equivalent to the target warming treatment for many studies, and 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).”
- iv. Lines 169-172, in the “Experimental infrastructure alters microclimate” section, where we write “These were the only studies in the MC3E database that monitored climate in 169 two types of control plots: structural controls (i.e., ‘shams’ or ‘disturbance controls,’ which contained all the warming infrastructure, such as soil cables (n=1), forced air chambers (n=2), or both (n=2), but with no heat applied) and ambient controls with no infrastructure added.”
- v. Lines 228-231 in the “Indirect and feedback effects of climate change manipulations ” section, where we write “Across these studies, all of which used infrared heating (two

with feedback control and one with constant warming), we found...”

We also detail the particular study types used in each analysis in the Supplemental Materials “Details of statistical analyses and results” section (e.g., page 1 under the “Analysis of effects of warming on daily temperature range” heading, where we write: “These data were from studies that used infrared warming with feedback control (n=2), infrared with constant control (n=2), forced air (n=2), and combined forced air and soil cable (n=2) warming types”), as well as in each supplemental table legend. In addition, to further emphasize to readers the importance of different warming treatment methodologies and the ways they may differ, we have added a new figure that shows photographs of the different warming methodologies included in the MC3E database (Figure 2), a new box (Box 1) that describes potential differences of the divergent warming methodologies included in the MC3E database, and a new table that separately summarizes warming, soil drying effects, and other indirect effects by warming technique and warming control, as the reviewer requests. The table includes quantitative summaries gleaned from the MC3E database, when possible, as well as qualitative statements from previous publications.

2. “Indirect effects”

- (a) *I think the authors still have a bias about indirect effects (their starting point is that they are somehow “anomalous” rather than part of the fabric of such experiments). The other initial reviewer who was even less positive than I was did not review the revised manuscript, giving the impression that the paper was now nearly ready for publication (i.e. the one of three most positive initial referees did give another ‘thumbs-up.’ the most negative referee did not chime in again and I was positive pending a more nuanced and careful presentation).*

We thank the reviewer for pointing out that our manuscript still appears to be biased about indirect effects. That is not our wish, as we agree with the reviewer that indirect effects are part of the fabric of such experiments—indeed, we feel that indirect effects are a critical component of the value of in situ active warming experiments. We rather wish to highlight the need to more fully analyze and interpret these indirect effects, as a better understanding of them is necessary for improved forecasting of biological responses to climate change. We describe the changes we have made to address this concern in detail below, under items b-d.

- (b) *The second of the two main points made in the Abstract (lines 10-13, three sentences) is about indirect effects. Here they have changed the wording to be more neutral in the first sentence, as recommended. The following sentence is poorly worded: “The implications of these complexities can have important biological consequences.” The combination of direct and indirect effects can be complex and can have important biological consequences (that would be one way to phrase this), but “implications” don’t have “consequences” (they more or less are consequences of consequences). The third sentence gives the one sensational example from their paper and it is stated as “one such consequence.” It is fair to talk about this one example but it is not fair to leave the impression that such effects might be the norm. How likely is it that such effects occur? The main text talks about this somewhat. I think the Abstract needs to put this one extreme example in a broader context.*

We thank the reviewer for pointing out our poor choice of wording. We have reworded the sentence to read (Line 11): ‘The combination of these direct and indirect effects is complex, and may have important biological consequences. We show a case study of plant phenology...’

We agree with the reviewer that it would be valuable to know how common indirect effects, such as the one we observed with budburst date, are. Indirect effects of this magnitude may be widespread, or they may not be very common—however, given the paucity of published

data across many studies we cannot infer the rareness or commonness of such effects. Indeed, this is one of the main research areas we are suggesting that researchers examine further. One of the goals of the current manuscript is to inspire additional investigation of other potential indirect effects, through data-sharing and future meta-analysis.

(c) *Regarding lines 311-319 and in particular this line:*

“Thus, researchers should not assume that the soil drying observed in warming experiments is likely to occur at all sites with future warming.”

The authors miss the point almost entirely here. Soil drying is likely to occur with future warming (it is largely a result of simple physics unless air humidity increases dramatically). The problematic issue regarding this point is with rainfall treatments (within warming experiments) not the warming treatments themselves. Let’s use a hypothetical example to illustrate. Imagine one set up a factorial experiment with three rainfall treatments and two warming treatments. Rainfall treatments are less-than-current ambient, current ambient, and more-than-current ambient, to test what happens with climate warming if predictions about future rain hold (i.e., whichever of the three treatments best match predictions for the place in question) and what happens if future rainfall is different than expected. Under all three rainfall scenarios and treatments, warmed plots would have higher ET and lower soil moisture than ambient temperature plots. Whether in the real world, soils in the future with joint changed temperatures and rainfall will be moister or drier than under current conditions hinges on both of those changes (temperature and rainfall). Our ability to predict direction of temperature change (it will be higher) is good, as is the impact of warming (experimental or real world) on soil moisture (higher ET will reduce it). The big unknown is whether and when and how much it will rain. So yes researchers should assume that soil drying as observed (at least directionally) in warming experiments is likely to occur, but whether those will be ameliorated or exacerbated by changing rainfall is what will determine future trajectories in soil moisture. [Note, increased total atmospheric vapor pressure would somewhat reduce VPG in a warmer world, but projections are for similar [ranging from slightly less to slightly more] total vapor pressure in growing seasons in the future, certainly not enough to eliminate increased ET].

We thank the reviewer for clarifying the point that his/her main concern in the previous review was the ability to forecast precipitation accurately. We added text to remind readers of the uncertainty around precipitation forecasts (Lines 323-325), where we now say: “For example, precipitation forecasts with climate change are uncertain, as are future changes in soil moisture. Soil drying in conjunction with future warming is forecasted in some regions...”

The reviewer further states that, all else being equal, warming should always lead to drier soils unless precipitation or humidity increases. While this may broadly be true, all else is rarely equal. The amount of drying depends not only on the atmosphere, but also land surface processes, including vegetation. As an example, see Figure 3 in Cook et al 2018. This figure shows, for an ensemble of climate models, the change in different moisture-related variables (soil moisture, runoff, precipitation) for the end of the 21st century under a high warming greenhouse gas scenario (RCP 8.5). While soil moisture does dry in most places, there are many areas (e.g., Australia), where the change is ambiguous (indicated by the hatching). This is because soil moisture changes in response to warming and precipitation changes, but it also changes with the *seasonality* of warming and precipitation changes (e.g., we don’t expect precipitation to change the same way in different seasons) and it depends on how the vegetation responds. To the latter point, the data included in MC3E are missing a key component of climate change: increased ambient CO₂ levels, which can have effects on plant physiology. One thing increased CO₂ does is increase water use

efficiency of plants and decrease stomatal conductance, which would reduce water losses and potentially compensate for warming induced increases in evapotranspiration. An additional point to consider is that the experimental infrastructure itself, at least for forced air and soil warming (the only warming types that conducted both structural and ambient controls) altered soil moisture, without adding heat. We have added text to clarify the complexities of forecasting soil moisture with climate change, and now cite this paper in Lines 330 and 335.

Whether the soil drying that occurs in experiments is reasonable or not, the main point we hope that readers come away with is that soil moisture should be quantified and interpreted in active-warming studies. This will allow temperature effects to be isolated and interpreted separately, and will improve our mechanistic understanding of effects of climate change on biological responses. We have tried to highlight this point in Lines 328-332, where we now say:

“Thus, the soil drying observed in warming experiments may not occur at all sites with future warming. The uncertainty associated with forecasting changes to soil moisture makes replicating future water availability regimes in climate change experiments especially challenging; one way to meet this challenge and make predictions—even given high uncertainty—is to quantify soil moisture effects in climate change experiments.”

- (d) *I don't think that indirect effects are a good analogy to Huston's hidden effects. First, there is nothing hidden about them. Second, his arguments were that they were something different for which species richness stood in as a proxy. That is not the same thing at all here.*

We thank the reviewer for pointing out the poor fit of this analogy. We have removed the reference to Huston's hidden treatments so that the phrase now reads: “These indirect effects are likely to have biological implications for many of the responses studied in warming experiments...” (Lines 272–274)

Summary: Because of what I view as inadequacy of the way both of the the main issues are handled, the paper (if printed in current form) will be sometimes cited as evidence for generic weaknesses in warming experiments and as evidence that to date global change experimental scientists have ignored indirect effects (which are painted as largely artifacts) in drawing conclusions about such studies. Instead these data should be used to highlight potential issues which occur more often in some types of warming experiments (i.e. tell us which technologies) than others. The authors still fail to be up-front about this, which is possible without formal statistics, merely by looking at the techniques and the available data.

We thank the reviewer for bringing up this issue. With this revision, we have modified the manuscript to be upfront about the important differences across methodologies: we have added a new box to highlight differences across different technologies, as well as mentioning these differences throughout the main text, as described above.

If the paper were printed in its current form, I think it will do more harm than good, because most readers won't have the background or take the time to try to wade through the complexities. I think it is the job of the authors and the journal to get this right. Finally, I should have said this earlier, but it occurred to me in reading through the paper quickly that the authors should provide some framework and context for their conclusions in light of differences among experiments and techniques. among plots within experiments and impacts of warming structures on microenvironment (two main examples of their focus), or that some experiments heat only soil, others heat only aboveground (with modest and seasonally variable impacts on soils), and some both? Such considerations are totally ignored.

The reviewer brings up an interesting and important question about comparing variation among plots within experiments versus impacts of warming infrastructure. The relative importance of their effects

is likely to vary by site. For example, if a site is spatially heterogeneous, for example with vegetation cover or soil type, this variation may interact with warming treatments to have a large effect on results. Infrastructure effects may be comparably small in magnitude at such sites; however, infrastructure may have a greater relative effect than spatial variation at homogeneous sites. We now discuss these points in the new Box 1, where we say “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 infrastructure shading may be lower).” We also write in the conclusion (Lines 365-368): “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.”

Our community too easily just throws together all warming experiments without such considerations, and in evaluating the questions of interest in their study, the authors do that here as well. Given their focus on drawing readers' attention to the potential problems with warming experiments it would be quite helpful to also frame this within the context of how we should think about experiments that warm just soils, just plants, or both. Sorry to not be able to be wildly supportive of this manuscript in its step-wise evolution; if I were not convinced that there was value in what they are trying to do, I would not have recommended continued path on Ecology Letters, but “potential” and “realized potential” are not the same thing.

We would like to take this opportunity to again thank the reviewer for the large investment of time and effort that has been dedicated to improving this manuscript. We greatly appreciate the attention to detail and thought-provoking questions that have been raised in an effort to enable this manuscript to realize its full potential, and have positive impact on the field of climate change ecological research. We have added a new box and table, with the aim of providing readers with some background on the complexities of different warming methodologies.

EDITOR

We have received responses from the two reviewers. Reviewer 2 remains convinced that the manuscript is ready for publication and will make an important contribution to the literature and the global change ecology field. Reviewer 1 also believes that the manuscript has potential to make a significant contribution to the field but remains concerned with specific aspects of the manuscript that could lead readers to incorrect and unjustified conclusions. Reviewer 1 offers detailed guidance for the authors to help them overcome the remaining issues (attached). The Reviewer's recommendations aim to clarify and thus strengthen the message of the paper. In particular the authors must ensure that conclusions are drawn with respect to the specific techniques and avoid drawing general conclusions where the results are driven by the limits of particular techniques (e.g. constant wattage). The authors should also heed Reviewer 1's suggestion to place the current work within a framework of current climate change experimental approaches (within the Discussion), and to draw on insights from other study sites that were not able to be included in the data analysis. I recommend major revisions in order to allow the authors time to fully address the issues raised by Reviewer 1.

We thank the editor for continued help in improving this manuscript. As detailed above, we have modified the figures and main text to include information about the specific techniques used, and we have added a new Box dedicated to this important topic. We have also added a new figure showing different warming techniques, and a new table, in which we cite 7 other studies.

REFERENCES

Cook, B.I., Mankin, J.S. Anchukaitis, K.J. (2018). Climate change and drought: From past to future.

Current Climate Change Reports, 4, 164-179.