



1300 Centre Street
Boston, MA, 20131

December 11, 2018

Dear Editors:

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.R2. We have incorporated the suggestions of the referees and editor by modifying the text, as detailed in the enclosed point-by-point response.

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 made six suggestions to reword portions of the text and Reviewer 2 suggested adding the magnitude of variation as a percentage, in addition to raw change, in the abstract. The handling editor provided a helpful summary of how we might address the suggested changes, and we have followed the editor’s suggestions. Please see our detailed, point-by-point response below for more information on the changes. Note that addressing comments by the reviewers made the abstract quite long, so we cut some text from the abstract in order to accommodate the further revisions requested by the reviewers (and we can adjust further, if requested).

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

I appreciate the steps taken by the authors to modify their paper and I sincerely hope that they view this as helping improve the paper and not merely as assuaging what they might see as an excessively opinionated and picky reviewer. The paper is nearly ready to be accepted, but I do find a few places where, in my view, the authors still seem to want to tell the story they began with, rather than the one the data provide. The authors are inconsistent in one instance they are happy extrapolating from a single case study to the world and yet within the data at hand they insist on using statistical significance (from a data base with very low power) to mask important differences in their data.

We are sorry the reviewer feels we were inconsistent, as this was not our goal. We have aimed to be consistent, in particular by focusing on significant results throughout the manuscript and using all available data. We use a single case study of budburst phenology as an example of how measured microclimate data, rather than target warming treatment, *may* affect our interpretation of temperature sensitivity, as these effects (of soil moisture and measured warming on budburst) are statistically significant. In contrast differences between feedback and control designs were not significant and thus we focused less on them. Further, we always used all available relevant data from the database in all analyses (i.e., the phenology case study analysis includes 5 studies—all the studies with soil moisture, above-ground temperature, and budburst phenology data). We have tried to be consistent and true to the data and analytical results throughout the manuscript.

Here are six points where I think language still needs modification.

POINT 1. In their response letter they write: “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 4.8° 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.”

And based on this view, they retained the following sentence in the abstract “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 varied in warming design and myriad ecosystem attributes.”

To me this is unacceptable. Given that basic physics predicts that constant wattage will be more variable and the raw data show this (“2.2 ° C variation on average for constant versus 1.1 ° C variation for feedback”), I believe it is misleading to lump them together in the Abstract. I have no problem with the authors noting both the lack of significance and extremely low sample size (which is not their fault) in the main text, but here I think it should read something like the following: “Our synthesis showed that measured mean warming, in plots with the same target warming, differed considerably; for example by 1.1° C, on average for experiments with feedback control and by 2.2° C on average for those using constant wattage.”

We have modified the above sentence to reflect the large variation in warming across studies. The warming differences that the reviewer mentions (1.1 versus 2.2) are for only a subset of warming types (infrared studies). We wanted to be clear about this, and wish to remind the reviewer and editor that there are other warming types included in our database (forced air and soil warming). Thus, we now say (Lines 8-11): “Our synthesis showed that measured mean warming, in plots with the same target

warming within a study, differed by up to 1.6°C (63% of target), on average, across six studies with blocked designs. Variation was high across sites and designs: for example, plots differed by 1.1°C (47% of target) on average, for infrared studies with feedback control (n=3) versus by 2.2°C (80 % of target) on average for infrared with constant wattage designs (n=2)."

POINT 2. Here is a case where in contrast they are fine with extrapolating from a single case study. "With a case study of plant phenology, we show how accounting for drier soils with warming triples the estimated sensitivity of budburst to temperature." As worded it sounds as if accounting for drier soils would routinely triple the sensitivity. I suggest that they re-word to make clear that in a single case study you found that accounting for drier soils in warming treatment triples the estimated sensitivity."

We thank the reviewer for pointing out a lack of clarity in our writing. To make it clear that the conclusion is drawn from one case study, we changed 'triples' to 'tripled'. We also added the number of experiments included in this case study, to be consistent across the abstract in highlighting the data underlying the numbers. The sentence now reads (Lines 13-14): "With a case study of plant phenology across five experiments in our database, we show how accounting for drier soils with warming tripled the estimated sensitivity of budburst to temperature."

Point 3. "We use a case study of spring plant phenology to demonstrate how analyses that assume a constant warming effect and do not include non-temperature effects of warming treatments on biological responses lead to inaccurate quantification of plant sensitivity to temperature shifts. " Again, this may or may not routinely "lead to inaccurate quantification of ?..". Re-word so it is clear you found it here; and that is a useful warming that it might occur sometimes or often in other studies.

We thank the reviewer for pointing out that, although we found an inaccurate quantification of plant sensitivity to temperature shifts in our case study, the presence and magnitude of such inaccuracies are unknown for other plant responses. To clarify this, we have added 'may' before 'lead to inaccurate'. The sentence now reads (Lines 67-69): "We use a case study of spring plant phenology to demonstrate how analyses that assume a constant warming effect and do not include non-temperature effects of warming treatments on biological responses may lead to inaccurate quantification of plant sensitivity to temperature shifts."

Point 4. Lines 143-151 "These studies include two infrared with feedback control, three infrared with constant wattage, and one soil warming cable with feedback control experiments. We found that the amount of observed warming frequently varied by more than 1C (mean= 1.6C, maximum = 4.9C) among blocks (Figure 3, 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). The differences in warming among blocks may be caused by fine-scale variation in vegetation, slope, aspect, soil type, or other factors that can alter wind or soil moisture, which in turn affect warming (Peterjohn et al., 1993; Kimball, 2005; Kimball et al., 2008; Hoepfner & Dukes, 2012; Rollinson & Kaye, 2015)." Here or in box 1, I think they need to add the paragraph from their letter (pasted above in Point 1) being more explicit about what they found.

While we appreciate the reviewer's concerns we feel this is already addressed in the paper. Though we do not use this exact text, the language and results of Box 1 are similar. Thus, we have made no changes to this section of the manuscript.

Point 5. Lines 326-341. Although improved, I think the text (in middles of paragraph and beyond) still implies that there can be a problem with warming-associated increases in ET and thus soil drying; this is unfortunate, because although warming under current rainfall may not predict soil moisture under future rainfall, the experimental problem is with inputs of moisture not with the way in which warmed plants and soils lose water! I think this needs to be made clear.

"In contrast, some responses documented in climate change experiments may not be in line with future

climate change?or may be too uncertain for robust prediction, and thus need explicit analyses and cautious interpretation. Although surface warming inevitably increases soil water evaporation, it does not necessarily translate to a decrease in soil water content. Precipitation forecasts with climate change are more uncertain than temperature forecasts, as are, consequently, future changes in soil moisture (Cook et al., 2018). For example, soil drying is forecasted in some regions, such as the southwestern United States, mainly because of reductions in precipitation and increased evaporative demand associated with warmer air (Dai, 2013; Seager et al., 2013). The northeastern United States, on the other hand, has been trending wetter over time (Shuman & Burrell, 2017), even though temperatures have warmed. Shifts in soil moisture are likely to vary by region, season, vegetation type, and soil depth (Seager et al., 2014; Berg et al., 2017; Cook et al., 2018). The uncertainty associated with forecasting changes to soil moisture makes replicating future water availability regimes in climate change experiments especially challenging; one way to meet this challenge and make predictions?even given high uncertainty?is to quantify soil moisture effects in climate change experiments. The altered light, wind, and herbivory patterns documented under experimental infrastructure (Kennedy, 1995; Moise & Henry, 2010; Wolkovich et al., 2012; Hoeppepner & Dukes, 2012; Clark et al., 2014b) represent other non-temperature effects that may be potential experimental artifacts and are worth quantifying in....”

We thank the reviewer for for sharing this feedback. In this section, our primary goal was to highlight the complications associated with forecasting soil moisture under climate change. We have made no changes to this section of the manuscript, as we feel the paragraph accurately presents some of the complications associated with predicting future soil moisture.

Point 6. Unless remedied elsewhere, a concern I had about the authors lack of holistic approach shows up again here regarding statements about belowground warming. They wrote “Plot shading and precipitation interference are likely to occur in chamber and infrared techniques, which both involve above-ground infrastructure, and less likely in methods that only warm from the soil. 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).” I think it makes sense to add “Of course, methods that warm only the soil and not aboveground completely, or almost completely, fail to warming the plants aboveground, which is a massively greater shortcoming compared to the advantage of not casting unintended shade.” As written it seems to imply that somehow ?soil only warming? is truly comparable to either aboveground or both soil and plant warming. To me, the difference is so massive that I think it does not really make sense to do so. I am not wild about how it is done here in the paper, but think it is okay for Ecology Letters as long as such statements are not made in a vacuum (i.e. don’t just note that lack of aboveground warming does not cast shade, without reminding readers of the far more important differences, like NO WARMING ABOVEGROUND).

We thank the reviewer for pointing out the need for a fuller discussion of the implications and trade-offs of different experimental infrastructure. We have added a sentence regarding the drawbacks of soil warming (Lines 638-640): “Soil warming methods, however, may be less representative of climate change, which will be driven by above-ground rather than below-ground warming. Warming cables also disturb the soil, potentially altering conductivity, water flow, and other soil properties. The biological impacts of such effects may be further enhanced or muted based on site characteristics (e.g., if a site is already heavily shaded, impacts from above-ground infrastructure shading may be lower).”

REVIEWER 2

I appreciate the detailed response to reviewer 3’s comments and the much more nuanced interpretation of the available data. I continue to see this manuscript to be a ‘call to caution’ in broadly generalizing results from global change experiments, something that we experimentalists all know but have not ourselves taken the time to quantify. I did not read this paper as general discouragement for doing these types of experiments; else we would all throw up our hands and give up. Rather it is a peer reviewed article that we can all cite when we acknowledge the limitations and complexities of the techniques we deploy. That said, it might resonate with a relatively small audience within the ecological community. I would hope it would not be used to discredit global change experiments, rather to improve the way

we conduct them and interpret the results. One sentence I still stumble on in the abstract is: “Our synthesis showed that measured mean warming, in plots with the same target warming, differed by 1.6°C, on average.... Can the authors change this to a percent change? That said, it is an improved manuscript. Kudos to the authors and to reviewer 3.

We greatly appreciate the time and effort the reviewer has dedicated to improving this manuscript. We now report both absolute and relative differences. The sentence now says (Lines 8-11): “Our synthesis showed that measured mean warming, in plots with the same target warming within a study, differed by up to 1.6°C (63% of target), on average, across six studies with blocked designs. Variation was high across sites and designs: for example, plots differed by 1.1°C (47% of target) on average, for infrared studies with feedback control (n=3) versus by 2.2°C (80 % of target) on average for infrared with constant wattage designs (n=2).”

EDITOR

The reviewers describe the recent revision as a high quality and important manuscript that quantifies – for the first time – the effectiveness of current climate change experiments in terms of their ability to manipulate micro-climate. It is an authoritative and informative review targeted at a broad audience of ecologists interested in species responses to climate change. The authors clearly demonstrate the consequences of smoothing over the inherent complexity of physical climate variables when using mean metrics to assess species responses to climate change, a practice that is common far beyond experimental ecology. The authors have responded comprehensively and positively to issues raised by one reviewer; the result is an excellent manuscript that will engage and inform a broad range of readers.

The reviewers request a few additional changes. Here I summarise how the authors might address these remaining issues.

We thank the editor very much for this helpful summary!

Reviewer 1.

Point 1. Separate the deviation from target warming between the two approaches.

We have made the suggested change, as discussed above in the response to Reviewer 1. The sentence now says (Lines 8-11): “Our synthesis showed that measured mean warming, in plots with the same target warming, differed by up to 1.6°C (63% of target), on average, across six studies with blocked designs. Variation was high across sites and designs: for example, plots differed by 1.1°C (47% of target) on average, for infrared studies with feedback control (n=3) versus by 2.2°C (80 % of target) on average for infrared with constant wattage designs (n=2).”

Point 2. To make it clear that the conclusion is drawn from one case study, change ‘triples’ to ‘tripled’.

We have made the suggested change (Line 14 in the Abstract).

Point 3. Add ‘might’ or ‘may’ before ‘lead to inaccurate...’

We have made the suggested change. The sentence now says (Lines 67-69): “We use a case study of spring plant phenology to demonstrate how analyses that assume a constant warming effect and do not include non-temperature effects of warming treatments on biological responses may lead to inaccurate quantification of plant sensitivity to temperature shifts.”

Point 4. The text reported in the authors’ cover letter is already included in Box 1, so no change is required.

We thank the editor for noting this, and have made no changes to this section of the manuscript.

Point 5. I disagree that the text in this section implies a problem with experimental approaches relating

to soil moisture but instead describes the difficulty in predicting likely future precipitation and thus expresses the inherent challenge of creating experimental soil conditions that are likely to be experienced in the future. Importantly, the authors have provided a workable solution, which is to report continuous metrics of soil conditions so that the effects can be modelled accordingly.

We thank the editor for this helpful feedback, and have made no changes to this section of the manuscript.

Point 6. Describe the tradeoffs faced by experimentalists between warming and shading effects caused by the respective methods.

We have added the following sentence to Box 1 (Lines 638-640: “Soil warming methods, however, may be less representative of climate change, which will be driven by above-ground rather than below-ground warming. Warming cables also disturb the soil, potentially altering conductivity, water flow, and other soil properties. The biological impacts of such effects may be further enhanced or muted based on site characteristics (e.g., if a site is already heavily shaded, impacts from above-ground infrastructure shading may be lower).”

Reviewer 2. The authors might consider reporting both absolute (degrees C) and relative (%) differences in temperature.

We thank the editor for this section. We now do report both absolute and relative differences. The sentence now says (Lines 8-11): “Our synthesis showed that measured mean warming, in plots with the same target warming, differed by up to 1.6°C (63% of target), on average, across six studies with blocked designs. Variation was high across sites and designs: for example, plots differed by 1.1°C (47% of target) on average, for infrared studies with feedback control (n=3) versus by 2.2°C (80 % of target) on average for infrared with constant wattage designs (n=2).”

We thank the editor for continued help in improving this manuscript.