The ARNOLD ARBORETUM of HARVARD UNIVERSITY

November 1, 2019

Dear Dr. Findlay:

Please consider our paper, entitled "Winter temperatures dominate spring phenological responses to warming" for publication as a "Letter" in *Nature Climate Change*. This manuscript is a revised version of an earlier submission (NCLIM-19081773). We include a point-by-point response to reviewer comments.

As you may recall, our manuscript utilizes an extensive new, extensive, global database to address a research topic of critical relevance to a broad swath of *Nature Climate Change* readers: the timing of spring phenology (e.g., budburst, leafout) in woody plants. Spring phenology impacts plant fitness, shapes plant and animal communities, and affects wide-ranging ecosystem services from crop productivity to carbon sequestration. Our work is groundbreaking in its synthesis of four decades of research across 72 experiments to quantify the relative importance of three environmental cues critical to phenology. We estimate overall chilling, forcing and photoperiod responses for 203 species from around the globe.

The three reviewers recognized the potential of our work to influence future research, as well as its interest to *Nature Climate Change* readers. They also highlighted some concerns. Reviewer 1 suggested that additional details and clarification of methods would be beneficial for a fuller evaluation of the study. Reviewer 2 felt unconvinced that the experimental methods synthesized in our meta-analysis could be reliably applied to natural systems. Reviewer 3 had reservations about the validity of the results given the data and modeling approaches used.

We have substantially modified the manuscript to address the concerns expressed by reviewers and the issues mentioned by the Editor after the initial submission. Specifically, we have added new text to the main manuscript, including methodological details, XX, and XX. We have also created a new figure, and modified previous figures in the main text to addres reviewer concerns. We have also added the online 'Methods' section to adhere to the new guidelines of *Nature Climate Change*.

Upon acceptance for publication, the database will be freely available at KNB (*?*; currently meta-data are there); the full database is available to reviewers and editors upon request. This work is a meta-analysis, so data have been previously published; however, the synthesis of these data and the tables, figures, models, and materials presented in this manuscript have not been previously published nor are they under consideration for publication elsewhere.

Sincerely,

Ailene K. Ettinger

Ailene Ettinger
Visiting Researcher, Arnold Arboretum of Harvard University Quantitative Ecologist, The Nature Conservancy- Washington Field Office

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

Reviewer 1 (Remarks to the Author)

The relative importance of forcing, chilling and photoperiod as cues for budburst is a fascinating one, with clear implications for predicting how species will respond to climate change. Here the authors leverage an exceptional dataset arising from experimental studies using sophisticated statistical analyses and arrive at the surprising conclusion that plants are generally more sensitive to chilling than forcing. I think this study has the potential to make a really valuable contribution that will be of broad interest to readers of this journal. However, I have quite a lot of criticisms/concerns of the study as it stands.

We thank the reviewer for the recognition that the OSPREE dataset is exceptional and that our study can make a valuable contribution to *Nature Climate Change* readers. We have revised the original manuscript substantially to address the reviewer's concerns, as detailed below.

(1) Models: The STAN modelling approach is sophisticated but I think the model is rather incomplete and this could affect the inferences that are reached. For instance why aren't terms included to allow the intercepts and slopes to vary across studies within species?

We completely agree with the reviewer that, ideally, our models would account for both variation in budburst responses among species and among studies. This was our aim in building the OSPREE database and using Bayesian models that can allow complex models to converge. During our data and model development, however, we found studies were generally confounded with species. This is a common problem in this area of research and other meta-analyses have faced similar issues (e.g., cite Kharouba PNAS). Without strong priors to help differentiate what variation should load onto the study versus species, it is difficult to fit both variables. Indeed, we evaluated models that included studies within species and studies crossed with species but both models were found to be unstable (e.g., estimates varied widely across model runs, model estimates had wide credible intervals, chains frequently did not converge).

To address this we worked to present a main model that only included species where we had multiple treatments across multiple studies. [Could add more here.]

We realized now we should have made these issues more clear. We now discuss this point on lines XX where we write 'Some species are often only represented in one dataset in the OSPREE database, making it impossible to statistically differentiate between species, study, and treatment effects for these taxa. To address this, we combined species found in only one study into "complexes" at the level of genera—such that each taxonomic unit we use in our model occurs across multiple studies (and treatments, see the *The Observed Spring Phenology Responses in Experimental Environments (OSPREE) database* section in the Supplemental Materials for details.'

Also, I would have thought there is very likely a geographic effect on the effects, and I suggest that you test whether the results are sensitive to inclusion of a spatial random term across which slopes and intercepts vary.

The reviewer makes an excellent point that budburst responses to temperature and photoperiod may vary due to the spatial location of studies or the geographic origin of plant material. In particular, budburst responses are expected to vary by latitude (11; 12; 13). We address variation in space by including a latitude model, which we now discuss on lines XX in the main manuscript. We focus on the spatial effect of latitude because there is strong evidence that there can be latitudinal differences in budburst responses, for example via interactions between latitude and photoperiod sensitivity (XX CITATIONS) and AND interactions with CHILLING?FORCING?) responses

Of course there may be other spatial factors besides latitude that affect responses to budburst. Indeed, there are numerous relevant questions that could be addressed with the OSPREE dataset. This is one reason that we are excited to publish the OSPREE database—we hope that these and other questions are explored by other researchers.

QUESTION: Do we want to add a model with continent (or UN units: Africa, Americas, Asia and the Pacific, Europe and Central Asia, and the Middle East) as a random effect? we could at least interpret the variation at this level.

(2) Meta-analysis: The analysis is described as a meta-analysis, but falls short of being a formal meta-analysis as it seems as though measurement error in the response variable is not incorporated. This should be straightforward to incorporate and I was surprised that it hadn't been given the complexity of the analyses. Also, please report the extent to which the approach followed recommendations made in the

PRISMA checklist

We thank the reviewer for pointing out that it would be helpful to know the extent to which our manuscript follows recommendations in the PRISMA checklist. We now have added an appendix reporting this, as well as the ROSE checklist, suggested b the *Nature Climate Change* Reviewer. . - currently in OSPREE 92% of rows of budburst data have 0, no response, nothing (vast majority) or "NA"- is this because we did not collect these data consistently or because they are not widely reported (my guess is both) - attach the dataframe to response .

(3) Methods: The methods seem to be missing from the main ms, and I kept flicking forward to consult a section that does not exist. I thought the Nature letter format does allow a methods section and I found it really to the detriment of the readability of the ms that there wasn't one.

We appreciate the reviewer's concern that methods are not easy to find. We were notified by the editors at Nature Climate Change that they will no longer be publishing Methods in print, but rather use a separate, online-only methods section. Therefore, all of our methods are now in a separate Methods section that will be available online, should our manuscript be accepted for publication in *Nature climate change*. We have added more more methodological information throughout the manuscript, however, to address this concern. For example, we added XX

(4) Chilling, forcing and photoperiod: In order for a reader to reach a conclusion about the robustness of the inferences it is vital that the method for quantifying cues is easily understandable. Currently in the main ms it is not (last paragraph of page 3). For instance, we are told the minimum temperature for chilling but not maximum, we are not informed as to when the chilling and forcing periods are and no discussion is given as to how the effect of photoperiod is modelled. It's also unclear in the main ms what a 'standard unit' (I see it is described in the supplement) is and this leaves the reader disconnected with what the analyses are doing. A simple remedy for this would be to include a schematic (as figure 1) that identifies the information used to quantify each cue and relate it to the response. In general the main ms does a very poor job of explaining what was done (the data used, how cues were inferred and vital details about what the models were estimating), instead referring the reader repeatedly to supplementary materials. While the supplementary materials are generally good I still felt disconnected from the data and how the cues were actually quantified. This could be addressed by taking some example datasets and working through in detail how the different metrics were calculated. Without knowing what was done I find it very hard to judge whether the main conclusions are robust.

-We appreciate the reviewers point...though this is challenging in a short-form paper. the methods are now fully described, including the upper and lower thresholds for chilling, in the online methods section (NEED TO ADD THIS)

(5) Chilling: I think it's important to know whether the inferences are robust to an alternative model of chilling, e.g., the sequential model that is widely used. From the supplementary materials it is clear that some effort has been made to consider alternatives (chilling portions) but given this analysis underlies the main conclusion of the paper I'd like to see alternative hypotheses considered.

We thank the reviewer for pointing out that our earlier version did not adequately compare results with the Utah model to results with the other chilling model we evaluated, Chill Portions. In the new version, we discuss this in lines 'XX where we write, "..." We have also added a supplemental figure showing chill portions (Figure SX) - compare this to Figures 2-3 (should we add this? not sure its necessary...) Ailene needs to look at this more. we do present both....perhaps 3d figures for both in supp?

(6) Estimates: It is surprising to see point estimates repeatedly reported throughout the ms without 95% credible intervals, this needs to be rectified. Also at present there is no formal test of whether the chilling response is significantly stronger than the forcing response, though this would be easy to do using the posteriors.

- we currently qhuse 50% credible intervals throughout- we can include these in more places - 95% credible intervals are in tables in Sup- can incorporate into main text more (Lizzie says use 96%) -Also add in response athat our approach is bayesian and cite some of Gelman (or some one elses's) stuff that 95% = arbitrary

Statistical artefact with linear regression (Page 5): That application of linear regression to data arising from a growing degree model can lead to biased estimates is a fascinating insight. However, in the supplementary materials it is not clear to me how the temperature sensitivity window for linear regression (for B. pendula or the simulations) is calculated/defined. How much can the issue of an advancing period of sensitivity be addressed by allowing the sliding window to shift over time? This issue is discussed in

Simmonds, E. G., Cole, E. F., & Sheldon, B. C. (2019). Cue identification in phenology: a case study of the predictive performance of current statistical tools. Journal of Animal Ecology.

-cite this paper somewhere in our manuscript. describe the window used (how many days before...I think it started sept 1) -sept 1. Dan and Cat are looking at effect of sliding window? - we agree this is an important issue. more research needed.

Minor comments

Page 2. I suggest changing 'high unexplained variation across' to 'substantial variation among'. We thank the reviewer for this suggestion and have made the recommended change (now Line XX).

Page 3. All three cues are not generally correlated in longitudinal studies; photoperiod and forcing are, but neither is usually very correlated with chilling.

We thank the reviewer for pointing out that our writing was not clear in this section. In our earlier version of the manuscript, we mentioned correlations between cues but did not specify clearly whether we meant correlations across space or time, nor were we clear about the scale or window at which these cues can be correlated. Though the reviewer states that chilling is often not correlated longitudinally, we have found that it can be correlated (Cite figure in supplement?). In addition, chilling and forcing are frequently negatively correlated in space (sites that have experience high chilling tend to have low forcing). In this new version of the manuscript, we have clarified our writing, which now says (Lines XX): " XX"

Page 3. Last sentence of paragraph 2. This is hyperbole. The mean is not expected to shift far beyond historical bounds, though the extremes clearly will.

The reviewer appears to have concerns about the following phrase (now lines XX): "... continued warming pushes climate into environmental regimes far beyond historical bounds." We thank the reviewer for highlighting this phrase, for which we clearly should have included citations to support. We now cite the IPCC forecasts that climate change is expected to push temperatures to XXX well beyond the historical bounds. We have added the following citations the the manuscript

Xu, Y., Ramanathan, V. and Victor, D.G., 2018. Global warming will happen faster than we think.

Page 3. Fourth line from bottom. Is interactions the correct term? i think so...what else could we say.."unquantified interactive effects" we think this is the correct term- we've tried to clarify the sentence.

Reviewer 2 (Remarks to the Author):

Spring leaf-out phenology plays a key role in terrestrial carbon and water flux, but the underlying processes are still unclear, especially how the environmental cues, including chilling, photoperiod and spring warm temperatures, interact and determine the leaf-out processes is still unclear, although most of the phenologist agreed that these three cues are all important. Therefore, quantify the relative importance are valuable and might be important for the phenology modeling and dynamics vegetation models. I carefully read this meta-analysis and found this is an interesting study, but I'm wondering, given the results were reliable, whether the meta-analysis results across experimental studies could reflect the natural plants' response? Or could we rely these experimental results that may inaccurate reflect underlying mechanisms? Because, according to the author (E.M. Wolkovich) previous study, the phenology under warming experiments could not reflect the natural observations (Wolkovich et al, 2012 nature, warming experiments underpredict plant phenological responses to climate change), which might arise from complex interactions among multiple drivers and remediable artefacts in the experiments that result in lower irradiance and drier soils.

We agree with the reviewer that there are many ways in which experimental conditions differ from observational conditions and as the reviewer points out, there are many differences between them. we highlight this with our chilling comparison- experiments use constant chilling vs natural conditions were temp. -what else to say here- perhaps a references that responses are correlated in experiments and natural world? (is there a primack or zonner paper that does this?) - call out for citations -foundational work on chiling- - the whole idea of this process is an established thing. - this are established methods developed because they work.

Furthermore, I'm not convinced that the chilling overweight forcing, and the effect of chilling, photoperiod and forcing might be quantified across more than 200 species based on the various manipulative experiments and MCMC-based Bayesian method, especially considering most of these experimental studies conducted only one year or less than 3 years. The main reasons come from: 1) most of these experi-

mental studies conducted with very different settings, such as using saplings vs. mature tree?s cuttings, how the ontogenetic effects play a role or impacts the results? Arbitrary controls in lights/photoperiod length/intensity vs. greenhouse natural light; in addition, for many experimental studies, the temperature and photoperiod were set under extreme climates. I would say this is a response to extreme climate. All of these factors might substantially affect the results.

2) the interact between chilling, photoperiod and forcing is complicated, and there are still unclear in many important facts. For example, the temperature thresholds of chilling and forcing estimation, and its species-specific values, are largely unknown. For some boreal or alps plants, they may budburst even when air temperature around freezing points, but the temperate trees are still dormancy even air T ¿ 15 degree; the correlations between eco- and endo-dormancy, corresponding the chilling and forcing, whether they are a parallel or a sequential pattern between chilling and forcing? When/how the photoperiod plays its role during the two phase dormancy? Once the endo-dormancy break, continuous chilling accumulation, for example a cold span during spring, is still active? Or entirely depending on the forcing? All these questions are still not figured out; 3) except chilling, forcing and photoperiod, other cues are also involved with the leaf-out processes, for example air humidity, see Laube et al, 2014 (but recently, Zohner et al, 2019 New phytologist deny this effect) and soil moisture and snow cover. Under manipulative conditions, these effect might be largely ignored as argued in Wolkovich et al, 2012 as well.

-we agree that much is unknown. this is one motivation for our paper. we hope that it encourages researchers to come up with new questions and approaches. Indeed, a study comparing fidnings in exp vs observational studies would be a useful addition that is beyond the scope of this paper.

4) species-specific response to chilling, photoperiod and forcing. This has been well reported, for example the pioneer species are opportunistic and photoperiod-insensitive, in contrast the late successional species are sensitive to photoperiod and higher forcing requirements, see the papers, as the authors cited, Krner & Basler 2010;2014; Laube et al, 2014; Zohner et al, 2016 and other studies. Across so large dataset/many species, the mean values, for example chilling effect is 2 times larger than forcing and photoperiod as well as its sensitivity, hold large uncertainty and are no sense.

One of the main conclusions is that chilling is over-weight forcing and recent advanced leaf-out is mainly associated with spring warming. However, this is inconsistent with recent study that found the spring phenology did not significantly change during the global warming hiatus, see its figure 1 in Wang et al, 2019 Nature comm, but the spring T is still significantly increase and winter getting colder over the Eurasian (Li, Stevens and Marotzke 2015 GRL)). It seems that increasing chilling and forcing could not explain the dynamics in spring phenology? How to explain this inconsistency?

- It seems that the reviewer is saying.....IfNot sure exactly what the reivewier's point is- that winters got cooler so we would expect spring phenology to advance if our hypothesis is correct? This is not necessarily true because we find that wamring is increaing chilling in mayn locations. Thus, cooling might be decreasing chilling- still could be consistent with the Wang et al paper. this actually may be consistent with our findings.

Minor commons Line numbers are needed;

We thank the reviewer for this suggestion and have added line numbers.

In methods, the study yielded data from 72 studies across 39 yrs... this is misleading, because for many experimental studies, table S1, the data only for one year, and most less than 3 yrs. -add what? a phrase saying, 'with most studies lastuing 1 year" add a figure - could look at effect of material

More description is needed of Bayesian hierarchical model in the main text; how to address this when no methods in main ms?!?

In the results sections, chilling has greater effect on budburst than forcing?. I would suggest providing the conditions, i.e. under future climate warming, due to the fact that these results come from experimental studies that simulated future warming, -sure- can say this ore carefully *In the results sections as well, the chilling only occur at warming above 4C? interesting, but does it occur across species? and locations?* -mean across species- explore variation? see maps in supplement

Zohner, Constantin M., et al. "Rising air humidity during spring does not trigger leaf?out timing in temperate woody plants." New Phytologist (2019). Wang, Xufeng, et al. "No trends in spring and autumn phenology during the global warming hiatus." Nature communications 10.1 (2019): 2389. Li, Chao, Bjorn Stevens, and Jochem Marotzke. "Eurasian winter cooling in the warming hiatus of 1998?2012." Geophysical Research Letters 42.19 (2015): 8131-8139. -cite these in main ms -could add humidity when to main text- *Reviewer 3 (Remarks to the Author):*

This manuscript addresses the relative importance of the environmental determinants of plant phenology using a meta-analytical approach. Specifically, the authors combine the experimental results of 72 studies and 203 species to estimate the effects of day length, winter chilling, and forcing on spring phenology, using hierarchical Bayesian models. The main finding is that almost all species respond to all three cues, with chilling having the largest, day length the smallest effect. Furthermore, the results suggest that, while all cues are important under experimental conditions, spring forcing will remain the dominant driver of spring phenology over the coming decades. The manuscript is well written and addresses a clear question. However, I have reservations as to the overall importance and validity of these results. That chilling is more important than day length has been shown by previous multi-species studies addressing this (e.g., Laube et al. 2014, Zohner et al. 2016). -true! other studies have looked at this. this is a meta-analytic approach. also, non of these studies attemted to look at 3 cues as tthe reviewer points outBut forcing is usually considered to be the most important!

Furthermore, the model output seems to suggest that all three cues (day length, chilling, and forcing) affect phenology in almost all species, leading the authors to conclude that their results contrast with the extensive literature [Zohner et al. 2016, Körner & Basler 2010] suggesting photoperiod is an unimportant cue for many species. [page 4]? Yet, when looking at Table S2, most of the species-level data they use are taken from Zohner et al. (2016) [Zohner16 database]. In fact, 173 (85%) of the 203 species included in this study were already investigated in Zohner et al. (2016). Given that in Zohner et al. (2016), 112 (65%) out of 173 studied species did not react to daylength at all, it is surprising that day length is reported as a relevant, consistent cue across species. This makes me wonder whether their hierarchical Bayesian model is confounded (e.g., giving to much weight to certain species ?complexes?) and thus not suitable for exploring the relative importance of the different environmental drivers of spring phenology.*

-interesting point about species differences. this is why we focus on the but compare only really abundance species: betula, fagus, quercus and - not sure what the reviewer means when s/he says that "most of the species-level data" are from Zohner. The Zohner dataset comprised 864/7459 (11%) rows in OSPREE. ITs true that the zohner dataset includes 144 out of 203 species in the full OSPREE budburst database. However, we include in the main model interpreted and presented in the figures only ispecies that wer across multiple studies. - To make this more clear, we have added columsn to S2 which lists models included in? -many zohner species excluded frmo main bb model. in model that includes all species (Table S3)- estimated effect of photoperiod does weaken (cforcing estimate gets stronger, chilling is about the same) -add comparison between all species and single model -look at zohner methods- could analyze this alone? oucld be because we estimated. - add somehing about partial pooling.- the reviewer seems to not understand this.

Körner & Basler 2010 clearly is an inadequate reference here, please delete

In our original submission, we cited two papers by Körner & Basler 2010: a 'Perspective' in *Science* (Körner & Basler 2010, *Science* 327, 1461) and a response to a critique of this perspective (Körner & Basler 2010, *Science* 329, 278). We believe that the reviewer is referring to the response and we agree that this reference is inadequte. We thank the reviewer for this comment, and have removed the reference from our manuscript.

Apart from that, I take issue with the estimation of the importance of forcing and the attempt to estimate the relative importance of day length, chilling, and forcing. First, I don't see how the effect of forcing can be disentangled from the effects of chilling. This would require knowledge on which temperature ranges are adequate to satisfy chilling and forcing requirements. Yet, as correctly stated in the Supplementary information (page 2), current models of chilling are hypotheses and likely to be inaccurate for many species. Similarly, the effective temperature ranges to fulfill forcing requirements are not known. As such, when comparing the relative importance of winter chilling versus spring warming both factors are likely to be confounded. Also, if a study uses two different forcing temperatures that both lie within the range of optimal forcing conditions, one would see no effect between the treatments and the authors would thus infer that forcing didn?t affect phenology, when in fact, forcing has a huge effect, not detected by the study design. Given these considerations, I don't think that a multivariate model, such as the one presented in this study, can adequately disentangle the relative importance of the three main phenological cues.

-great point!impossible to disentangel. much more info needed at species level. in absence of this...what is approach? one motivation for this paper is to highlight the need for additional work. -we rely on the original researchers to separate forcing from chilling conditions- these were the treatments that they imposed. we therefore assume that they used a range of treatments that are relevant for their focal species. -the reviewer does not suggest an alternative approach....

-cite new schematic figure - one of the findings of our paper- is that we need more work to accurately separate chilling and forcing - the perceptive reviewer has highlihgted a general problem of the field...and yet...our models has been exrtemely predictive. Supplementary material

p.2: What do you mean by ?we included only studies with at least 49.5% budburst?? This is not correct for most of the studies included in your OSPREE dataset. E.g., Heide (1993) and Zohner et al. (2016) defined budburst as the date when 1?3 buds on a twig had opened. Please clarify. - we have reworded the section to clarify. it now says ... ‘’

p. 3: Total chilling ranged from -1304 to 4724 Utah units? The Utah model allows for negative chilling units? What?s the biological justification for that? -the biological justification, as described by the developer of the original model (Richardson) is ... - It has been suggested that Utah model is not useful in subtropical areas (), however ,the vast majority of studies in our metanlyses were conducted in temperate areas (XX/XX). in addition, utah units were used by the majority of studies included in our analyses (62/740 with temperature data making up 92.8 % of the 7643 rows of the OSPREE database that measured days to budburst). We used the Utah model to report chilling because that allowed us to include the greatest amount of studies from the OSPREE database (i.e., because many studies used this model to estimate chilling).

p. 4: Latitude model: This model doesn't make sense to me. What is the latitude you refer to here? The location where the experiment took place? You refer to provenance locations, I doubt these are available for most of the studies, especially the ones conducted in botanical gardens or other collections.

We thank the reviewer for highlighting the need for greater clarity and methodological details about this model. In the initial submission, we did not clarify what was meant by latitude. The latitude we refer to is provenance latitude (the latitude from which the plant source material was collected). We were unable to identify provenance latitude for only one study (Theilges et al 1975), or 10/2882 rows of data. We have made the following changes to address this reviewer's question:

1. Lines XXX in the main text: we now say ‘While photoperiod had the smallest effect among the three cues, our results contrast with the extensive literature suggesting photoperiod is an unimportant cue for many species (Zohner et al 2016,Körner & Basler 2010)—instead we found it was surprisingly large, even when accounting for its interaction with provenance latitude (i.e., the latitude from which plant material was collected; see Supplemental Materials for details...)
2. Supplemental Materials, Page 4: in the description of the Latitude Model, we now say: 'we examined the effect of including latitude in a model similar to our main one, but designed to estimate latitude effects. This model estimated the effects of each phenological cue (chilling, forcing, photoperiod) on days to budburst (as in the main model), in addition to the effect of provenance latitude (i..e., the latitude from which plant material was collected) and the interaction of photoperiod by provenance latitude. We include this interaction because photoperiod effects are expected to vary by latitude...' and later in this paragraph we now say ‘..then subsetted the species and species complexes to include only those that had multiple provenance locations across different latitudes.'
3. Supplemental Materials, Caption for Table S5: We have replaced 'latitude' with 'provenance latitude' so that the caption now reads: ‘Using a model with Utah chilling units and testing the effects of provenance latitude plus the interaction between provenance latitude and photoperiod results in slightly muted effects...'
4. Supplemental Materials, Caption for Figure S3: We have replaced 'latitude' with 'provenance latitude' so that the title of the caption now reads: ‘Estimates for effects of chilling exceeded estimates for forcing, photoperiod, provenance latitude, and the interaction between provenance latitude and photoperiod, for most species...'

Figures: Figs. 2 and 3, showing a 3-dimensional illustration of the interplay between winter chilling and spring warming, are very hard to read. I would prefer a simpler illustration.

We appreciate the Reviewer's perspective. We have shown both simpler 2-dimensional versions of these figures and the more complex 3-dimensional versions to a number of scientists and found opinions to be split on preferences for 2d versus 3d versions. To address this reviewer's concern, we have moved the 3-dimensional version of the manuscript to the supplemental materials, and now include 2 dimensional versions of Figure 2 and 3 in the main text.

References cited 11. J. Gauzere, et al., Agricultural and Forest Meteorology 244, 9 (2017). 12. K.

Saikkonen, et al., Nature Climate Change 2, 239 (2012). 13.D. A. Way, R. A. Montgomery, Plant, Cell & Environment 38, 1725 (2015).