

*Reviewer comments are in italics.* Author responses are in plain text.

## **Reviewer #1 (Remarks to the Author)**

*The authors have done a very good job of addressing my comments and I really enjoyed reading this version. I now just have a couple of further quite minor suggestions.*

We thank the reviewer for these kind words, as well as his time and thoughtful suggestions for improvement throughout the review process.

*1) The explanation of chilling is greatly improved. I still feel a bit uncomfortable as to whether this really is capturing chilling (or whether the form that chilling takes for peach is really appropriate for these high latitude species), but I accept that this may be as much as can be achieved with these data. I was pleased to see that the assumptions/limitations are much more clearly presented. Is it possible for each study to identify the days over which forcing and chilling units were accumulated? Perhaps this could then be presented as a figure. For instance, if days before event was the x axis and study/population was up the y axis – with different colour used to indicate the accumulation of forcing and chilling units. It would help the reader to get a sense of the degree to which chilling units precede or coincide with forcing units.*

We have added a new figure (new Figure S1), as the reviewer suggested, to the Supplemental Information. This figure presents the days over which chilling and forcing were accumulated for each row of data in our analyses.

*2) I thank the authors for checking the results are robust to within subject mean centering. I think this analysis is worth adding to the supplementary materials as it will help to convince readers who might otherwise have been sceptical of the results like myself. While I agree with the authors' argument in the response to reviewers that purely observational studies may be more susceptible to having spatial differences in mean conditions, such differences are plausible in experimental studies too - for instance, at higher latitudes one might very likely adopt a lower range of temperatures. Fig S2 is a great addition.*

We have added the table summarizing the within subject mean centering approach to our Supplemental Materials (Table S7), as the reviewer suggested.

*3) Latitude:photoperiod interaction - This interaction is tested but rather inexplicably only the fact that it is significant is mentioned, with no discussion of the effect. From table S9 it looks as though as you go further north there is a non-significant tendency for the photoperiod effect to weaken, which I think is pretty interesting and consistent with what one might expect. I wonder (no need to follow this up) if the chilling effect also interacts with latitude as one might also expect.*

We thank the reviewer for pointing this out. Because the interactive effect was not strong (i.e., 50% uncertainty intervals overlap zero), we do not mention the direction of the interaction.

### *Minor comments*

*Line 9. When you make the comparison of the magnitude of the effect of the different cues this must all be very dependent on the scale across which they have been manipulated. Some brief comment to that effect would aid interpretation.*

We thank the reviewer for this suggestion. The abstract has been modified extensively by the Editor, and this comment is no longer relevant.

*Line 22. I think this sentence would benefit from rewording to make the meaning clearer.*

We have reworded the sentence in question to improve clarity in our intended meaning. The sentence now says: "Though responses to warming are widespread, showing strong advances on average, there is substantial

variation among species and sites.”

*Line 115. The fact that the credible intervals overlap could simply arise if parameters are poorly estimated. A better measure would be the % difference of the effect sizes. In table S5 the two chilling methods give very similar estimates, but the model including all species gives estimates that are more strikingly different.*

We thank the reviewer for pointing this out. Indeed, the percent difference between the effect sizes is quite low, as well. We have modified the text to include this. It now says: “While we found that applying a different chilling model did not strongly affect our estimates (*i.e.*, 95% uncertainty intervals of estimates for chilling, photoperiod, and forcing effects overlapped using two different chill metrics, Utah and chill portions, and the mean posterior of these estimates varied by about 10% or less between the two metrics, see Table S5)...”

*Congratulations on a fascinating and very detailed study.*

Thank you!

*Signed*

*Ally Phillimore*

*[I sign all of my reviews]*

## **Reviewer #2 (Comments for the Author):**

*Thanks for the reply. The estimation in chilling and forcing is still challenge, which I totally agree with the reviewer 1, although the gdd or chilling days were widely used in phenology studies. The temperature thresholds should be site- and species-specific, as well as the chilling periods should be spatial different even within species, and thus the estimation on the relatively weight of these cues, among these experimental studies, likely involve larger uncertainty, even if the statistic method is novel and reliable.*

The reviewer points out an ongoing challenge of estimating forcing and chilling, when the extent of site- and taxa-specific variation in temperature thresholds for these cues are poorly understood. It is our hope that our manuscript and the full OSPREE database will shine a light on this challenge and inspire additional research to address it. We now mention the potential for site- and species-specific temperature thresholds in chilling in Lines 107-110, where write: “Thus, while researchers generally define “chilling” and “forcing” treatments based on temperatures in controlled experiments (including in the studies synthesized here, see Fig. 1), fully separating out what plants experience as chilling versus forcing (as well as how this varies across species and sites) will likely require new methods to measure endo- and ecodormancy...”

*Furthermore, the experimental results, should carefully applied in the natural conditions because the experimental setting is largely different from the natural conditions, and the temperature response might be substantially different as well. I would thus suggest the authors to highlight the results carefully, but highlight the results comes from a meta experiments.*

We thank the reviewer for this important reminder. We now highlight that our findings come from a meta-analysis of experiments in the following places:

- Abstract: “Here we use a global meta-analysis of all published experiments to test the relative effects of these cues.”
- Line 49: “Here, we leverage these short-term controlled environment experiments in a meta-analysis...”
- Lines 145-147: “Reinterpreting our estimates of effects of chilling, forcing, and photoperiod (from experiments) using climate and phenology data that have led to observations of declining temperature sensitivities in Central Europe suggests that chilling and photoperiod are unlikely to cause the observed declines.”

- Figure 3 caption: “Estimates of budburst across a range of forcing temperatures and estimated chilling (converted to a representative mean temperature, see *Estimating chilling* in Methods) based on overall estimates of chilling and forcing effects from a meta-analysis of short-term experiments using controlled temperature and/or photoperiod conditions”
- Figure 4 caption : “Here we show species-level estimates from our model based on a meta-analysis of experiments...”