

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

Associate Editor's comments to the authors

We thank the authors for their patience with the reviews. Two reviewers provide several suggestions to improve the manuscript. I will suggest authors to particularly improve the clarity on their recommendations including their feasibility as suggested by the first reviewer. The second reviewer has raised some methodological concerns, which need careful consideration in the revision.

We are grateful to the Associate Editor and two Reviewers for their comments and feedback on our initial submission. With your input, we feel that we have much improved this manuscript. In particular, we have restructured our recommendations for addressing the problem of periodicity to address the feasibility issues highlighted by Reviewer 1 and the broader methodological applications pointed out by Reviewer 2. These changes are detailed below.

Reviewer: 1

The manuscript evaluates different experimental designs for the investigation of interactive effects of temperature and light on spring phenology of plants. The reported results and conclusions may also apply to other fields of research where interactions of these two factors are tested. The article deals with a fundamental question in ecological research and has great implications for the design of experimental studies and the interpretation of data coming from such studies. The arguments are straightforward and the examples mentioned are clearly presented and comprehensible. The assessment of the mathematical calculations leading to Fig. 3 are beyond my expertise, nevertheless the results seem plausible and easy to follow.

We thank the Reviewer for their thoughtful comments on our manuscript and we are glad they found the topic timely and important.

I appreciate the importance of the topic and definitely acknowledge the need to remind the experimental community that the choice of experimental design has far reaching implications for the outcome of the studies and that a lot of effort has to be put in designing experiments - especially if resources are limited! Having said that, I'm wondering though whether the conclusions of the study will ever take effect in the experimental community as the proposed designs reach a level of complexity that can only be realised with extreme effort and very high costs. It seems that the authors have come to a similar conclusion, which left me somewhat perplexed.

We agree with the Reviewer in their assessment that if the only solutions to this problem involve substantial increases in experimental costs and effort, it will be difficult for the experimental community to adopt our suggestions widely, and we can see how the Reviewer would think this based on the way we presented the solutions in our original submission. Our actual view is that there are a number of options for addressing the issues we described in the paper, ranging from simple statistical corrections and re-interpretation of experimental effect sizes to the admittedly more complicated and costly experimental re-designs that the Reviewer has flagged.

To better demonstrate the range of options available to researchers, we have substantially restructured our presentation of the solutions, especially our "Paths Forward" section, to emphasize both the low-cost, easily adopted ways to account for the issues that arise through the experimental co-variation of thermo- and photo-periodicity along side of the more complicated experimental designs that we presented in our original submission (lines 75, 245-251 and 253-271).

In addition to providing a new exemplary statistical correction in 245-251, we also now state:

Fundamentally, simply recognizing the issues that arise when thermo- and photoperiods are experimentally co-varied and accounting for this in interpreting effect-sizes and reporting uncertainty is a powerful start for improving inference from experiments. This awareness can be applied both forward and backward; to future experiments that seek to understand the interactive effects of temperature and photoperiod and to synthesizing and interpreting the near-century's worth of research in this area that has already been published.

We believe that these changes will encourage experimentalists to take on this issue at whatever scale their time and resources allow, and are grateful for the Reviewer for highlighting this critical need.

Reviewer: 2

This study described the common problems of periodicity (a latent experimental covariation) in experimental manipulations for testing the interactive effects of temperature and light. The authors concluded by outlining several experimental designs that can overcome the problem of periodicity. The topic is important in experimental designs to mechanistically assess organismic responses to the environment. The control treatment was not considered in this review, and then generalized experimental designs that improve statistical orthogonality of controlled environment experiments could be further offered. In general, I think there are two major concerns that should be addressed. Therefore, I would like to recommend a substantial revision before I can recommend to be accepted for publication in Functional Ecology.

We thank the Reviewer for their time and feedback on this manuscript. We appreciate their perspective on our examples, and have made substantial changes to the way we present them, and to how we structure our “Paths Forward” section. We detail these changes below.

Major concerns:

1. The authors detailed the problem of inference that can arise when manipulating the periodicity of both temperature and light in experiments. These details are based on the two treatment levels of two variables, although authors emphasized that an experiment must have at minimum two treatment levels of at least two variables. However, an experiment that includes three treatment levels (e.g., cooling-control-warming) of at least two variables may be more realistic and common in the future. Therefore, I suggest that this paper should give clear details based on three treatment levels of two variables. At least, the experiment might include three levels of photoperiod treatments (16, 12, and 8 hours) and two levels of forcing treatment (30/20 and 20/10 °C day/night), and vice versa.

We agree with the Reviewer that many experiments benefit from having more than two levels of light and temperature treatments, especially in the context of predicting climate change responses, which is one of the major applications we address in the text.

We have addressed these concerns in two major ways. First, to highlight the critical relevance of experimental periodicity covariation to larger experiments with more than two treatment levels, we have adapted our text to explicitly make the connection between simple, two-level examples and more realistic multi-level experiments in several places (line 133, lines 276-281, line 291). Additionally, for our example illustrating how covariation of periodicity can be an issue for any photoperiod

study whether or not temperature is an explicit experimental treatment (line 166) we have added three levels of photoperiod to the text to help illustrate this point.

Second, to visually illustrate how the problem of periodicity scales with increasing number of treatment levels, we have generated a new figure for the Supporting Information that reproduces Fig. 2 with three levels of photoperiod and forcing (Fig. S1). We also added text that details this connection in line 187.

Because sub-fields of ecology may have different standards and capacities for the number of treatment levels included in experiments, we chose to provide examples with just two treatment levels for all variables to serve as a “minimum reproducible example” to clearly illustrate our conceptual and mathematical points. We think that this simplicity best illustrates the concepts of our paper, and feel that the changes to the paper that we have detailed above make it more clear how these simple examples can be applied more broadly to experiments of increasing complexity. If the Reviewer and editor still find that providing more examples with more than two treatment levels in the main text would help to clarify the issues presented in the paper, we would be happy to revisit this.

2. This manuscript presents several alternative experimental designs. However, these designs were concluded from the experiment with the two treatment levels of two variables. Moreover, the alternative experimental design, like “1. Manipulate photoperiod and temperature intensity with no thermoperiodicity”, sacrifices the biological realism of diurnal temperature variation. Therefore, I suggest that the experimental designs which balance biological realism with robust inference, experimental effort with statistical power, and account for the effects of unmanipulated or unmeasured variables should be offered; further, whether the experimental designs with various variables (>2) and treatment levels (>2) also could offer.

We appreciate the Reviewer’s feedback here, and in response, have re-framed our “Paths Forward” sections to offer a more comprehensive view of the options for overcoming the problems we present in the paper, emphasizing some may be more appropriate than others depending on the biological processes and organisms of interest to researchers and the specific goals of their experiments. In this version of the text, we have included more guidance, and new citations, about when simpler designs may be appropriate. For example, regarding the design that “manipulates photoperiod and temperature intensity with no thermoperiodicity”, we now clarify in line 291 that:

...many aspects of physiology and development do not appear to respond explicitly to diurnal temperature variation...so in many cases this experimental simplification may be worthwhile to improve inference on the overall individual and combined effect of temperature and photoperiod on biological processes.

Additionally, to further highlight the importance of larger experimental designs in these forward-thinking recommendations, we have added a paragraph in the “Paths Forward” section (lines 320-331) that both acknowledges “two level studies may struggle to estimate interactions due to under-sampled treatment levels” and highlights that experiments with many levels are one of the most powerful approaches: “for detecting interactions and non-linearities among variables.”

We feel these additions will greatly improve readers’ ability to translate the concepts we lay out in this paper to their particular sub-fields and experimental systems, and thank the Reviewer for pushing us to solidify this connection.

Minor comments:

L 121-127: please provide detailed information about the controlled environment experiments

We have updated the information in this section and added a reference to a recently published paper detailing the patterns of experimental treatment applications in the controlled environment experiments of the OSPREE database in line 124.

L 132-139: please give the same example on photoperiod and forcing treatments with Fig. 4 and other parts of this paper.

We thank the Reviewer for noticing this inconsistency and have adjusted all figures to match the in-text examples.

L 179-180: why regression models will always underestimate the forcing effect?

In the updated manuscript, we explain the reason for this in line 184 stating: “This is because forcing is the variable with latent, unmeasured variation.”

L 208-210: please provide more detail about the difference between the two experimental designs.

In this version, we have included details about the treatment levels in each experiment in the “Modeling Methods” section of the Supporting Information. In the main text we have added a reference to this section at line 217 to highlight this.

L 241: intensity not intesity

Thank you for catching this typo. We have corrected this in the updated manuscript.

Fig. 4b-c: the line where the temperature rises should be vertical, not diagonal.

Thanks, we have made this change.

Fig. 4c: please give more detail about why need these experimental efforts and what is the representation of each experimental effort.

We have adapted the main text (lines 312-319) to elaborate on this design. There we state:

Such experiments are the best way to assess the comparative importance of temperature intensity and periodicity, which would provide important insights towards parsing the relative strength of temperature and light cues and their interactions, (Fig. 4c). However, executing such an experiment with a full-factorial design would substantially increase the size of a study and its experimental effort and, given the availability of the statistical tools and simpler experimental designs we discuss above, may only be worthwhile when researchers are explicitly interested in the relative contributions of periodicity and intensity in a particular study system.

We have also updated the figure’s caption to reflect this information as well.