

*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 been able to much improve this manuscript. In particular, we have restructured our recommendations for addressing the problem of periodicity to address the feasibility issues highlighted by R1 and the broader methodological applications pointed out by R2. 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.*

*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 thank the Reviewer for their thoughtful comments on our manuscript and glad they found the topic timely and important. We agree with their assessment that if the only solutions to this problem involve substantial increases in experimental costs and effort, the takeaways from our study will be difficult for the experimental community to adopt widely. However, we see 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 highlighted. In order to better present this range of options we have substantially restructured our "Paths Forward" section to emphasize both the low cost, easily adopted ways to account for the issues that arise through the experimental covariation of thermo- and photo-periodicity along side of the more complicated experimental designs that we presented in our original submission (lines 244-250 and 252-263). We believe that this change will encourage experimentals 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 thorough feedback on this manuscript. We appreciate their perspective on our examples, and have made substantial changes to the way we present our examples and 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, and that such designs may become more common as technology improves. We chose to provide examples with just two treatment levels for all variable to serve as a “minimum reproducible example” to illustrate our conceptual and mathematical points as well as allow for the most direct comparison to our case study on real data. With that said, within the framework of linear regression addressed in this paper, we appreciate the Reviewer’s point that the issues we present in this paper apply to studies with any number of treatment levels, and we have tried to highlight this fact directly in line 132, line 193, and lines 269-274. 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 165) we have added three levels of photoperiod to the text to help illustrate this point.

We hope that the Reviewer finds the *minimum reproducible example* justification of our framework compelling and that the many explicit references to studies with more than two treatment levels we have added make it more clear about how to relate the issues we discuss in this paper to larger experiments with many treatment levels. If the Reviewer still feels strongly that providing examples with more than two treatment levels in the main text would help to clarify the issues presented in the paper in some way, we would be happy to further discuss the pros and cons of this approach.

*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 wish there was a single, clear design that always achieved the goal of “balancing biological realism with robust inference, experimental effort with statistical power, and account for the effects of unmanipulated or unmeasured variables”. Instead, we believe there are a range of solutions, that vary in complexity and effort, of which 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. As we indicated above, we have restructured our “Paths Forward” section to try and serve as a better guide for these tradeoffs.

Additionally, the Reviewers suggestion to address experiments with more than two variables in our “Paths Forward” section is well taken. While we maintain that examples with two experimental treatments each with two treatment levels are the most clear and parsimonious way to present the points we make in the paper, the Reviewer is correct that our suggestions can, and should, be applied for experiments with any number of treatments and levels. For example, the experimental design depicted in Fig. c is an example of an experiment with three variables (temperature intensity, thermo-periodicity and photo-periodicity). To further highlight this point, we have added an additional introductory paragraph in the “Path Forward” section to emphasize that our generalized design can be applied to experiments with any number of treatment levels and any number of variables (lines 269-274).

*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 125.

*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 183.

*L 208-210: please provide more detail about the difference between the two experimental designs.* In this version, we have included detail about the treatment levels in each experiment in the “Modeling Methods” section of the the Supporting information (lines ??-??). In the main text we have added a reference to this section at line 216 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 for this suggestion, 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 (line 308) to elaborate on this design and updated the caption as well.