Experimental designs for testing the interactive effects of temperature and light in ecology: the problem of periodicity

D.M. Buonaiuto [1,2,a], M. Donahue[3], E.M. Wolkovich[1,2,4]

Author affiliations:

[1]Arnold Arboretum of Harvard University, Boston, Massachusetts, USA. ORCID: 0000-0003-4022-2591

[2]Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, Massachusetts, USA

[3]Hawai'i Institute of Marine Biology, University of Hawai'i at Manoa, Kan'eohe, HI, USA.

[4]Forest & Conservation Sciences, Faculty of Forestry, University of British Columbia, Vancouver, British Columbia, Canada

[a]Corresponding author: 617.823.0687; dbuonaiuto@g.harvard.edu

Abstract

1. Temperature and light cues interact to control many biological processes. Experiments give researchers the ability to manipulate these environmental cues independently, and can be designed to robustly quantify their individual and interactive effects on any particular biological activity. Testing the interactive effects of multiple environmental cues require experimental treatments to be fully independent, and any unmeasured experimental covariation among treatments can result in incorrect conclusions regarding cues effects.

2. Using a database of controlled environment experiments on the spring phenology of woody plants as a case study, we highlight how a common experimental set-up, designed to parse the interactive effects of temperature and photoperiod on time to budburst, introduces a latent experimental covariation of these treatments by coupling photo- and thermo- periodicity. Using data simulations, algebraic corrections and a comparative analysis of published experiments, we demonstrate how this unmeasured experimental covariation biases statistical inference regarding the relative contribution of light and temperature cues to phenological variation.

3. We identify this experimental covariation in more than 40% of published phenology studies that manipulate photoperiod. Our analyses demonstrate that the coupling of thermo- and photo- periodicity results in the overestimation of the effect of photoperiod, the underestimation of forcing effects, and misleading conclusions about their interactions on phenology. This may, in part, explain why the significance of photoperiod cues for spring phenology is currently debated in the literature.

4. Accurate forecasting of how varying environmental conditions will impact the dynamics of biological events requires the accurate quantification of cues responses. To this end, we present several alternative experimental designs that can provide more robust estimations of the relative effects of temperature and photoperiod on phenology, and many other biological processes controlled by temperature and light.

Keywords: forcing, full-factorial, growth chamber, light, phenology, photoperiod, temperature, thermoperiod

2942 words, 4 figures

Introduction

Across the tree of life, temperature and light availability shape a number of important biological processes including growth and metabolic rates (MacLean & Gilchrist, 2019), sex determination (Brown *et al.*, 2014), acclimatization to seasonal environments (Hamilton *et al.*, 2016) and the timing of life cycle transitions (i.e., phenology, Forrest & Miller-Rushing, 2010). These biological responses in turn dictate broad-scale ecological processes and patterns ranging from biogeochemical cycling (Piao *et al.*, 2007) to species range limits (Chuine & Beaubien, 2001). Characterizing the specific dynamics of how these environmental factors synergistically affect biological processes across a wide range of taxa has become even more important as anthropogenic global change continues to expose organisms to novel environmental conditions (Pörtner & Farrell, 2008).

Because temperature and light availability often co-vary in the field (for example, in most temperate ecosystems, daylength and temperature both increase as the season progresses, Rosenberg, 1974), it can be difficult to disentangle their relative contributions to biological processes. In contrast, experimental manipulations of climate variables in artificial environments can mechanistically characterize biological responses to environmental fluctuations (Ettinger *et al.*, 2020; Primack *et al.*, 2015). Researchers have used controlled environments of all shapes and sizes to this end (Downs, 1980), and these efforts have greatly advanced our collective understanding of the fundamental biology of a wide variety of organisms and ability to predict ecological and evolutionary responses to current and future climate change (Stewart *et al.*, 2013).

However, controlled environment experiments have their own challenges. Experimentalists must balance biological realism with statistical inference, experimental effort with statistical power, and account for the effects of unmanipulated or unmeasured variables (Scheiner & Gurevitch, 2001). Because biological responses to the environment are generally the product of complex interactions between multiple environmental signals (Casal, 2002), seemingly small choices about experimental designs can generate significant differences in outcomes. Experimental treatments are rarely standardized among researchers, even within disciplines (Wolkovich *et al.*, in Review.), and these complexities may in part contribute to the many discrepancies between experimental studies and observation data (Poorter *et al.*, 2016). Even with these limitations, controlled environment studies remain a powerful tool for mechanistically assessing organismic responses to the environment, provided that the implications of treatment designs are well understood and well matched with the scope of the research question.

As technology advances and experiments become more complex, researchers can manipulate more variables and multiple axes of variation (e.g., temperature, amplitude, periodicity, wavelength) at the same time. Yet these efforts may present a tradeoff between biological realism and statistical inference. Through investigating the literature on experiments with plant phenology, we show that experiments that manipulate both photo- and thermo- periodicity often introduce a latent experimental covariation between light and temperature treatments, which may misrepresent the effects of each of these environmental variables and the interaction between them. We begin by briefly detailing how temperature and light treatments are generally applied in phenology experiments and review the minimum experimental elements required to robustly test interactions between two or more environmental variables. We then detail the problem of inference that can arise when manipulating the periodicity of both temperature and light in experiments, and demonstrate the extent to which this is an issue through data simulations, a mathematical correction, and a comparative analysis of published experiments. Finally, we conclude by outlining several experimental designs that can overcome the problem of periodicity. While, our example deals with phenology of temperate woody plants, the issues and solutions we present below are broadly applicable to studies on any other organisms and biological processes that utilize temperature and light signals.

Estimating phenological cues from experiments

Decades of experimental work in controlled environments have demonstrated that temperature (both cool temperatures in fall/winter and warming temperatures in spring) and photoperiod are the primary phenological cues for plants in the temperate/boreal zones (Ettinger *et al.*, 2020). While exposure to cool winter temperatures (chilling) strongly impacts phenology (Laube *et al.*, 2014), we focus here on warm temperature and light treatments, because controlled chilling treatments with light are uncommon (Wolkovich *et al.*, in Review.). Choices about how to apply warm temperature and light treatments, in particular, can compromise inference on their effects, so we focus on these two cues.

While a large variety of experimental designs have been used to study plant phenology, generally experiments tend to manipulate two major axes of light and warm temperature variation:

1. Intensity: The amount or quality of a variable. Here we define temperature intensity as the amount of heat present in the system (measured in degrees). In the phenology literature this measurement is generally referred to as forcing. We define light intensity as the luminosity or irradiance present in the system (measured in lumens or watts).

2. Periodicity: The interval at which the intensity of the variable is applied. Hereafter, we refer to the periodicity of light as photoperiod (often used synonomously with "daylength") and the periodicity of temperature as thermoperiod.

For phenology, photoperiodicity is generally considered the primary light cue for plants (Way & Montgomery, 2015), (though regarding light intensity and phenology see Brelsford & Robson, 2018; Cober *et al.*, 1996). For temperature, conventionally both intensity and periodicity drive phenological activity and several metrics (e.g. growing degree hours, thermal sums, growing degree days) that combine these two axes have been developed (Gu, 2016). The importance of thermo-intensity and periodicity is well supported; under natural conditions diurnal temperature fluctuations in temperate regions can be quite large in the spring, and studies have found that diurnal temperature variation strongly influences plant phenology (Burghardt *et al.*, 2016). In fact, even if thermoperiodicity is not an explicit treatment variable (i.e., manipulated systematically), incorporating it in experiments can be essential for translating experimental results into real world predictions (Chiang *et al.*, 2020).

Like many other biological processes, recent advances have demonstrated that plant phenological responses are nonlinear, due largely to interactions between cues (Wolkovich *et al.*, in Review.; Fu *et al.*, 2015), highlighting the need for experiments designed to evaluate the strength of these interactions. To have the statistical power to partition the individual and interactive effects of two or more variables, an experiment must:

1. Have at minimum of two treatment levels of at least two variables.
2. Treatment levels must be full factorial (Fig. 1a.). Full factorial designs are both balanced (Fig. 1b.) and orthogonal (Fig. 1c.); meaning that all possible treatment combinations are applied and each treatment is independent of all others (Cheng, 2016).

These two critical elements may seem obvious but are conspicuously absent from many published studies. Using a recently published database of woody plant phenological experiments, OSPREE: Observed Spring Phenological Responses in Experimental Environments (Wolkovich *et al.*, 2019), we found that out of 152 controlled environment experiments (across 93 studies) only 18 manipulated both light and forcing cues with a design that was both balanced and orthogonal. But even experiments that are designed to be full factorial frequently violate the assumption of orthogonality when both photo- and thermo- periodicity are built into experiments. We detail this problem below.

The problem of periodicity

A common approach in phenology experiments that seems to balance prior knowledge about the underlying physiology of phenology, biological realism and experimental inference is to vary photoperiodicity, and thermal intensity and periodicity (e.g., Flynn & Wolkovich, 2018; Sanz-Perez *et al.*, 2009; Basler & Körner, 2014). For example, a basic experiment might include a long (16 hours) and short (8 hours) photoperiod treatment and a high (25/15°C day/night) and low (20/10°C day/night) forcing treatment. In this case, the thermoperiodicity is not an explicit treatment (both high and low temperature treatments use a diurnal fluctuation of 10 °C), and is simply incorporated in the design to enhance biological realism. At first glance, this design appears to meet the criteria of a full factorial design, multiple treatment levels that are balanced and orthogonal, with high/low temperature treatments (mean 20°C and 15°C respectively) and long/short photoperiod treatments applied in all possible combinations.

Yet the orthogonality of this design is based on the assumption of a 12 hour thermoperiod. If, rather the thermoperiod is coupled with the photoperiod, the temperature treatment is non-orthoginal because the daily mean temperature of the long/high treatment will be higher than that of the short/high treatment, and the long/low treatment slightly warmer than the short/low. We refer to this experimental set-up as a coupled design (i.e. thermoperiod and photoperiod are coupled with each other). Coupled designs introduce an experimental covariation between photoperiod and forcing treatments. This experimental covariation is clearly illustrated when temperature treatment levels are converted to thermals sums. We calculate thermal sums (also called growing degree hours), by multiplying hourly temperatures above a certain base temperature threshold by the number of hours for which they are applied over a 24 hour period (Parent *et al.*, 2019). For example, given a base temperature of 0°C, a low forcing treatment of 20/10°C day/night accrues 400 thermal units per 24 hours when crossed with the long (16 hour) photoperiod treatment and only 320 thermal units when crossed with the short (8 hour) photoperiod treatment. While this experimental covariation among the photoperiod and temperature treatments is biologically realistic, it makes it statistically impossible to differentiate the independent and interactive effects of temperature and photoperiod on any given biological process.

This problem of inference that arises from the experimental covariation of thermo- and photoperiodicity is not limited only to studies seeking to directly compare the effects of photoperiod and forcing; it applies in any study evaluating the influence of photoperiod on biological activity, even if it is the only manipulated cue. Experimentally isolating the effect of photoperiod assumes that all other environmental variables are held constant. Similar to the case described above, the coupling of photoperiod and thermoperiod in an experiment where forcing is supposed to be a consistent, background condition (e.g., two levels of photoperiod treatments of 8 and 16 hours, both at a background temperature of 20/10°C day/night) would yield a situation in which longer photoperiod treatments were also receiving more, unmeasured heating than shorter photoperiod treatments. In this case, some amount of the perceived photoperiod effect is due to the latent, increased forcing, and the experment will not isolate the true effect of photoperiod.

Of the 51 experiments in the OSPREE database that manipulated photoperiod experimentally, up to 43% of them appear to include an experimental covariation with thermoperiod. Of the 18 studies that manipulated both photoperiod and temperature interactively, we found that up to 55% of them appear to have this issue, suggesting that the true interactive effects of these cues on spring phenology is quite poorly characterized. This may be in part why the relative contribution of temperature and photoperiod cues to spring phenology remains a contentious debate in the phenology literature (Koerner & Basler, 2010; Chuine *et al.*, 2010; Körner & Basler, 2010).

Periodicity and inference

If the lack of orthogonality introduced to experiments when photoperiod and thermoperiod are coupled is overlooked, regression models will always overestimate the photoperiod effect and underestimate the forcing effect on spring phenology (Fig. 2a,b.). This is because forcing is the variable with latent, unmeasured variation. In the case of phenology, this is particularly significant because studies repeatedly suggest that forcing is a more dominant cue than photoperiod for spring phenology (Chuine *et al.*, 2010; Zohner *et al.*, 2016; Gauzere *et al.*, 2019).

If experiments are designed to quantify the interaction between photoperiod and forcing, here too, the experimental covariation of periodicity will result in an erroneous estimation of the interaction. (Fig. 2c,d.). Our simulation depicts a particularly troublesome case where a true sub-additive interaction is interpreted as a supra-additive one (Fig. 2c,d.), however, this must not always be the case. Experimental covariation of light and temperature treatments due to coupling thermo- and photo- periodicity will generally result in the incorrect estimation of the interaction term, but the exact nature of this statistical issue depends on the sign and strength of the interaction.

We can attempt to estimate how much of a photoperiod effect is due to forcing in experiments where they covary by making several major assumptions. Perhaps the biggest assumption we make is that forcing and photoperiod effects are additive and linear (i.e., there is no interaction). While this may not be true in nature, it gives us insight into the potential effect of the experimental covariation of periodcity by allowing us to solve algebraically for the separate effects of forcing and photoperiod. We replace the qualitative factor (high/low forcing) by the quantitative effect of forcing (thermal sums) to properly account for the difference in forcing between short and long photoperiods (see Supporting Information: Estimating the effects of experimental periodicity covariance mathmatically). Using the data from one experiment that experimentally coupled thermo- and photo-period, Flynn & Wolkovich (2018), we found that 33% of the published photoperiod effect of 4.5 days could be due to forcing.

We note that our algebraic solution cannot be as readily applied in experiments that assume photoperiod and forcing interact. However, we can generally assess the scope of the problem of inference due to experimental covariation of periodicity by comparing studies that used a coupled design to those with alternative approaches. While we are aware of no experiments that explicitly compare the effects of experimentally coupling vs. uncoupling photo- and thermo- periods, two phenology experiments in our lab utilized many overlapping treatment levels and species from the same sampling sites, however in one study, Flynn & Wolkovich (2018), photo- and thermo- period experimentally co-vary, while in the other, Buonaiuto & Wolkovich (2021), photo- and thermo- period were varied independently. Comparing the cue estimates from these two studies offers an opportunity to test our theoretical and mathematical predictions, and further understand the uncertainty in cue estimates due to coupled periodicities.

We subset each dataset to include only the species shared among the two studies, and re-analyzed the data using Bayesian hierarchical models to compare the difference in the photoperiod, forcing and interaction estimates (see Supporting Information: Modeling Methods). We found that the estimated differences in the mean response to photoperiod and forcing and their interactions among study designs were on the same order as our predictions. We estimated a substantially weaker (less negative) photoperiod effect, and marginally stronger forcing effect for the uncoupled vs. coupled experimental design (Fig. 3). The interaction term we estimated for the uncoupled design was negative, suggesting the interaction between photoperiod and forcing is supra-additive, while the estimated interactive effect from the coupled design was sub-additive (Fig. 3).

Unlike in our simulations (Fig. 2), in this comparison we cannot assess what the “true” effects of these variables are. There are almost certainly other factors driving the differences between these experiments. Both were conducted in different years, sampled different individuals from the population, and used different methods for applying chilling pre-treatments (Flynn & Wolkovich, 2018; Buonaiuto & Wolkovich, 2021). However, because this comparison is well matched to our predictions and prior knowledge about how temperature and photoperiod are expected to interacting in phenology, we argue that the influence of experimental periodicity covariation on statistical inference is apparent enough to take seriously.

Paths Forward

We have systematically demonstrated that experiments that couple thermoperiod and photoperiod cannot robustly differentiate the individual or interactive effects of temperature and photoperiod on spring phenology (or any other biological process) due to an unmeasured experimental covariation among temperature and light treatments. Given the paucity of interactive studies in the literature, it is clear that we need more well designed studies to better characterize the effects of these cues. Below we offer several generalized experimental designs that improve statistical orthogonality of controlled environment experiments, which could be further developed and adjusted to fit the needs of experimentalists across many sub-fields of ecology and evolutionary biology.

1. **Manipulate photoperiod and temperature intesity with no thermoperiodicity**. This approach allows for the maintenance of statistical orthogonality across treatment combinations (4a.). The main drawback is that this design sacrifices the biological realism of diurnal temperature variation, which may make it more difficult to translate estimates from experiments to real world applications.

2. **Compensitory diurnal temperature fluctions**. There are almost unlimited pairs of integers that can reduce to the same mean (e.g. $(24 + 26)/2 = (30 + 20)/2 = 25$) and the non-orthogonality of the mean daily temperature (or thermal sums) that arises in a coupled photo-thermo- period design could be corrected for by proportionately increasing the diurnal temperature fluctuation of the short photoperiod treatment relative to the long treatments (4b.). However, if the differences between day and night temperature has a meaningful biological effect, this introduces another confounding, non-orthogonal factor for interpreting temperature and photoperiod effects. For example, the influence of day time warming of phenology can be as much as three times stronger than proportionate night time warming (Rossi & Isabel, 2017; Meng *et al.*, 2020).

3. **Uncouple thermoperiod and photoperiod**. By varying thermoperiod and photoperiod independently (4c.), statistical orthogonality can be maintained across treatments. However, this approach may also introduce new artifacts that occur from the biological rather than statistical interactions between light and temperature (Chew *et al.*, 2012). For example, there is evidence that increasing temperatures in the first two hours of daylight can be almost as effective for stimulating shoot elongation as similar temperature increases for the whole photoperiod (Erwin, 1998). With this design, treatments must inherently differ in the amount of time the warmer daytime temperature extends into the dark nighttime light regime, introducing a new axes of non-orthogonality.

In correcting one problem, each of these designs introduces another, which may in fact be an intrinsic property of any experimental manipulation. It may also be that the experimental design that best balances environmental realism, statistical inference and translatability to observational studies are designs that continue to couple periodicity to mimic natural systems. Moving forward, researchers using this design need to be aware of its non-orthogonality, and carefully consider how to present the uncertainty around their effect estimates. It would be useful for researchers to explicitly test how cue estimates vary among experimental designs, and which design is most useful for predicting phenology in the field under current and future climate conditions. In the meantime, we hope that our presentation of this issue is a reminder that we experimentalists must continue to be thoughtful about matching our experimental designs to the goals of a study, and be transparent about uncertainty around our experimental inference.

Conflict of Interest Statement:

The authors declare no conflict of interest.

Author contributions

DMB, MD and EMW conceived of the manuscript; MD and EMW developed the algebraic solution; DMB performed the comparative analysis of the published studies; DMB led the writing of the manuscript. All authors contributed to writing and gave approval for the submission.

Data Availability

Data from the Flynn & Wolkovich (2018) study is available at the Harvard Forest Data Archieve (https://harvardforest1.fas.harvard.edu/exist/apps/datasets/showData.html?id=HF314) and from the Buonaiuto & Wolkovich (2021) study available at Knowledge Network for Biocomplexity (https://knb.ecoinformatics.org/view/doi:10.5063/PG1Q4B). The R code used to analyse the data is available on github.

References

Basler, D. & Körner, C. (2014) Photoperiod and temperature responses of bud swelling and bud burst in four temperate forest tree species. *Tree physiology* **34**, 377–388.

Brelsford, C.C. & Robson, T.M. (2018) Blue light advances bud burst in branches of three deciduous tree species under short-day conditions. *Trees-Structure and Function* **32**, 1157–1164.

Brown, E.E., Baumann, H. & Conover, D.O. (2014) Temperature and photoperiod effects on sex determination in a fish. *Journal of Experimental Marine Biology and Ecology* **461**, 39–43.

Buonaiuto, D.M. & Wolkovich, E.M. (2021) Differences between flower and leaf phenological responses to environmental variation drive shifts in spring phenological sequences of temperate woody plants. *Journal of Ecology* **109**, 2922–2933.

Burghardt, L.T., Runcie, D.E., Wilczek, A.M., Cooper, M.D., Roe, J.L., Welch, S.M. & Schmitt, J. (2016) Fluctuating, warm temperatures decrease the effect of a key floral repressor on flowering time in arabidopsis thaliana. *New Phytologist* **210**, 564–576.

Casal, J.J. (2002) Environmental cues affecting development. *Current Opinion in Plant Biology* **5**, 37–42.

Cheng, C.S. (2016) *Theory of Factorial Design*. Chapman and Hall/CRC.

Chew, Y.H., Wilczek, A.M., Williams, M., Welch, S.M., Schmitt, J. & Halliday, K.J. (2012) An augmented arabidopsis phenology model reveals seasonal temperature control of flowering time. *New Phytologist* **194**, 654–665.

Chiang, C., Bånkestad, D. & Hoch, G. (2020) Reaching natural growth: The significance of light and temperature fluctuations in plant performance in indoor growth facilities. *Plants* **9**.

Chuine, I. & Beaubien, E. (2001) Phenology is a major determinant of tree species range. *Ecology Letters* **4**, 500–510.

Chuine, I., Morin, X., Bugmann, H., Koerner, C. & Basler, D. (2010) Warming, photoperiods, and tree phenology. *Science* **329**, 277–278.

Cober, E., Tanner, J. & Voldeng, H. (1996) Soybean photoperiod-sensitivity loci respond differentially to light quality. *Crop Science* **36**, 606–610.

Downs, R.J. (1980) Phytotrons. *Botanical Review* **46**, 447–489.

Erwin, J. (1998) Temperature and light effects on stem elongation. *Journal of the Japanses Society For Horticultural Science* **67**, 1113–1120, Session of Plant Growth Regulation by Physical and Mechanical Stimuli at the Commemorative Symposium of the 75th Anniversary of Japanese-Society-for-Horticultural-Science, JAPAN, APR 03-04, 1998.

Ettinger, A.K., Chamberlain, C.J., Morales-Castilla, I., Buonaiuto, D.M., Flynn, D.F.B., Savas, T., Samaha, J.A. & Wolkovich, E.M. (2020) Winter temperatures predominate in spring phenological responses to warming. *Nature Climate Change* **10**, 1137–1142.

Flynn, D.F.B. & Wolkovich, E.M. (2018) Temperature and photoperiod drive spring phenology across all species in a temperate forest community. *New Phytologist* **219**, 1353–1362.

Forrest, J. & Miller-Rushing, A.J. (2010) Toward a synthetic understanding of the role of phenology in ecology and evolution. *Philosophical Transactions of the Royal Society of London B: Biological Sciences* **365**, 3101–3112.

Fu, Y.S.H., Zhao, H.F., Piao, S.L., Peaucelle, M., Peng, S.S., Zhou, G.Y., Ciais, P., Huang, M.T., Menzel, A., Uelas, J.P., Song, Y., Vitasse, Y., Zeng, Z.Z. & Janssens, I.A. (2015) Declining global warming effects on the phenology of spring leaf unfolding. *Nature* **526**, 104–107.

Gauzere, J., Lucas, C., Ronce, O., Davi, H. & Chuine, I. (2019) Sensitivity analysis of tree phenology models reveals increasing sensitivity of their predictions to winter chilling temperature and photoperiod with warming climate. *Ecological Modelling* **411**, 108805.

Gu, S. (2016) Growing degree hours - a simple, accurate, and precise protocol to approximate growing heat summation for grapevines. *International Journal of Biometeorology* **60**, 1123–1134.

Hamilton, J.A., El Kayal, W., Hart, A.T., Runcie, D.E., Arango-Velez, A. & Cooke, J.E. (2016) The joint influence of photoperiod and temperature during growth cessation and development of dormancy in white spruce (Picea glauca). *Tree Physiology* **36**, 1432–1448.

Koerner, C. & Basler, D. (2010) Phenology under global warming. *Science* **327**, 1461–1462.

Körner, C. & Basler, D. (2010) Response—warming, photoperiods, and tree phenology. *Science* **329**, 278–278.

Laube, J., Sparks, T.H., Estrella, N., Höfler, J., Ankerst, D.P. & Menzel, A. (2014) Chilling outweighs photoperiod in preventing precocious spring development. *Global Change Biology* **20**, 170–182.

MacLean, H.J. & Gilchrist, G.W. (2019) Temperature, photoperiod and life history traits in drosophila subobscura. *bioRxiv* p. 717967.

Meng, L., Zhou, Y., Li, X., Asrar, G.R., Mao, J., Wanamaker, A.D. & Wang, Y. (2020) Divergent responses of spring phenology to daytime and nighttime warming. *Agricultural and Forest Meteorology* **281**, 107832.

Parent, B., Millet, E.J. & Tardieu, F. (2019) The use of thermal time in plant studies has a sound theoretical basis provided that confounding effects are avoided. *Journal of Experimental Botany* **70**, 2359–2370.

Piao, S., Friedlingstein, P., Ciais, P., Viovy, N. & Demarty, J. (2007) Growing season extension and its impact on terrestrial carbon cycle in the northern hemisphere over the past 2 decades. *Global Biogeochemical Cycles* **21**.

Poorter, H., Fiorani, F., Pieruschka, R., Wojciechowski, T., van der Putten, W.H., Kleyer, M., Schurr, U. & Postma, J. (2016) Pampered inside, pestered outside? differences and similarities between plants growing in controlled conditions and in the field. *New Phytologist* **212**, 838–855.

Pörtner, H.O. & Farrell, A.P. (2008) Physiology and climate change. *Science* **322**, 690–692.

Primack, R.B., Laube, J., Gallinat, A.S. & Menzel, A. (2015) From observations to experiments in phenology research: investigating climate change impacts on trees and shrubs using dormant twigs. *Annals of Botany* **116**, 889–897.

Rosenberg, N.J. (1974) *Microclimate: the biological environment*. Wiley, New York.

Rossi, S. & Isabel, N. (2017) Bud break responds more strongly to daytime than night-time temperature under asymmetric experimental warming. *Global Change Biology* **23**, 446–454.

Sanz-Perez, V., Castro-Diez, P. & Valladares, F. (2009) Differential and interactive effects of temperature and photoperiod on budburst and carbon reserves in two co-occurring Mediterranean oaks. *Plant Biology* **11**, 142–51.

Scheiner, S.M. & Gurevitch, J. (2001) *Design and analysis of ecological experiments*. Oxford University Press.

Stewart, R.I.A., Dossena, M., Bohan, D.A., Jeppesen, E., Kordas, R.L., Ledger, M.E., Meerhoff, M., Moss, B., Mulder, C., Shurin, J.B., Suttle, B., Thompson, R., Trimmer, M. & Woodward, G. (2013) *Chapter Two - Mesocosm Experiments as a Tool for Ecological Climate-Change Research*, vol. 48, pp. 71–181. Academic Press.

Way, D.A. & Montgomery, R.A. (2015) Photoperiod constraints on tree phenology, performance and migration in a warming world. *Plant, Cell & Environment* **38**, 1725–1736.

Wolkovich, E., Chamberlin, C., Buonaiuto, D., Ettinger, A. & Morales-Castilla, I. (in Review.) Integrating experiments to predict interactive cue effects on spring phenology with warming .

Wolkovich, E.M., Ettinger, A.K., Flynn, D., Savas, T., Chamberlain, C., Buonaiuto, D. & Samaha, J. (2019) Observed spring phenology responses in experimental environments (OSPREE). doi:10.5063/F1CZ35KB.

Zohner, C.M., Benito, B.M., Svenning, J.C. & Renner, S.S. (2016) Day length unlikely to constrain climate-driven shifts in leaf-out times of northern woody plants. *Nature Climate Change* **6**, 1120–1123.

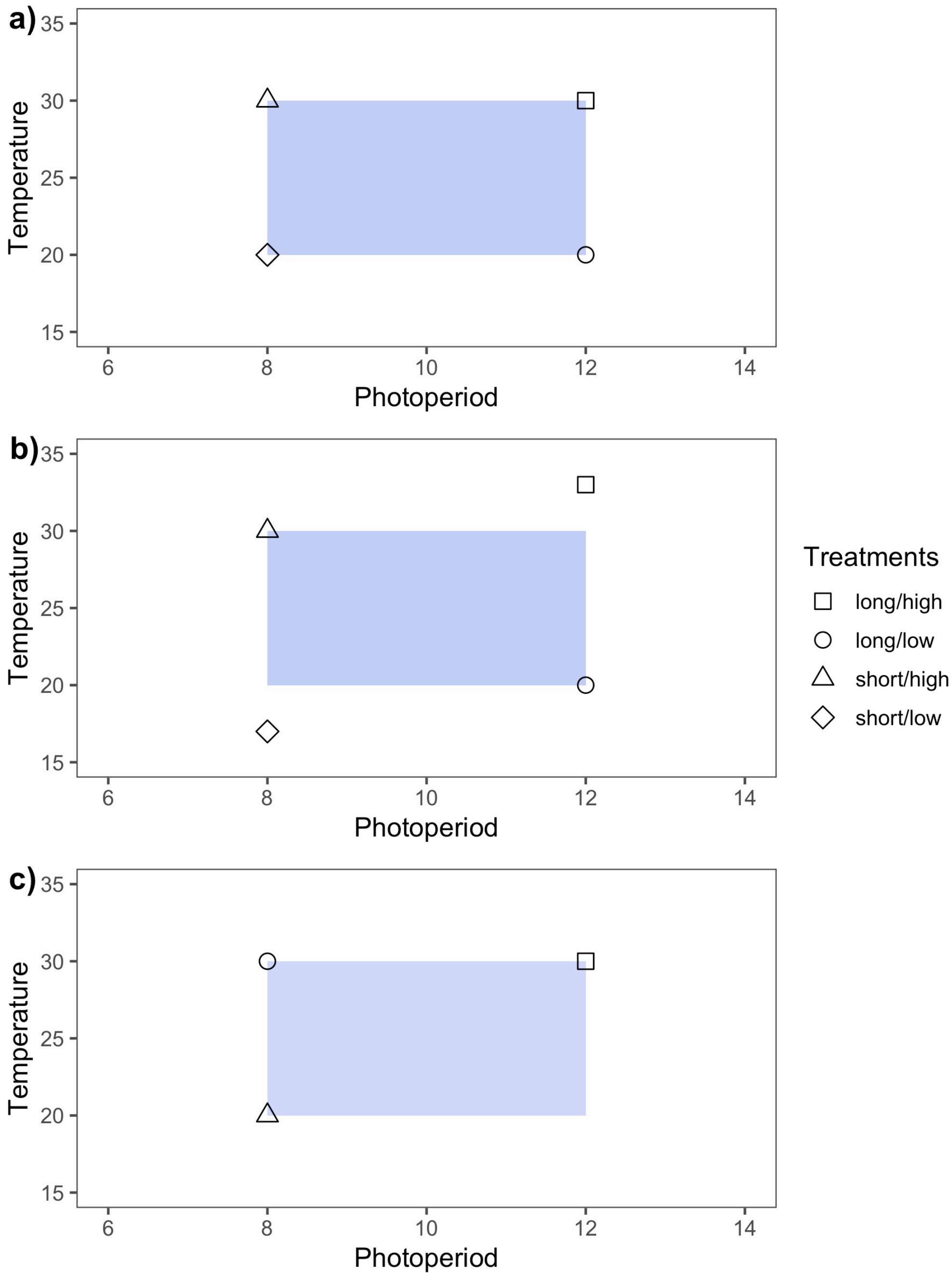

Figure 1: Idealized experimental designs demonstrate three approaches for varying temperature and light treatment levels in controlled environment experiments. Design **a)** is full factorial in that treatments levels are balanced and orthogonal. This design is appropriate for testing interactions between two or more variables. In **b)** the design is balanced but not orthogonal. Non-orthogonality in experiments can arise when experimental covariation among the manipulated variables is not accounted for. In **c)**, the experimental design is orthogonal but unbalanced. Lack of balance in experiments often arises due to time, space or resource limitations.

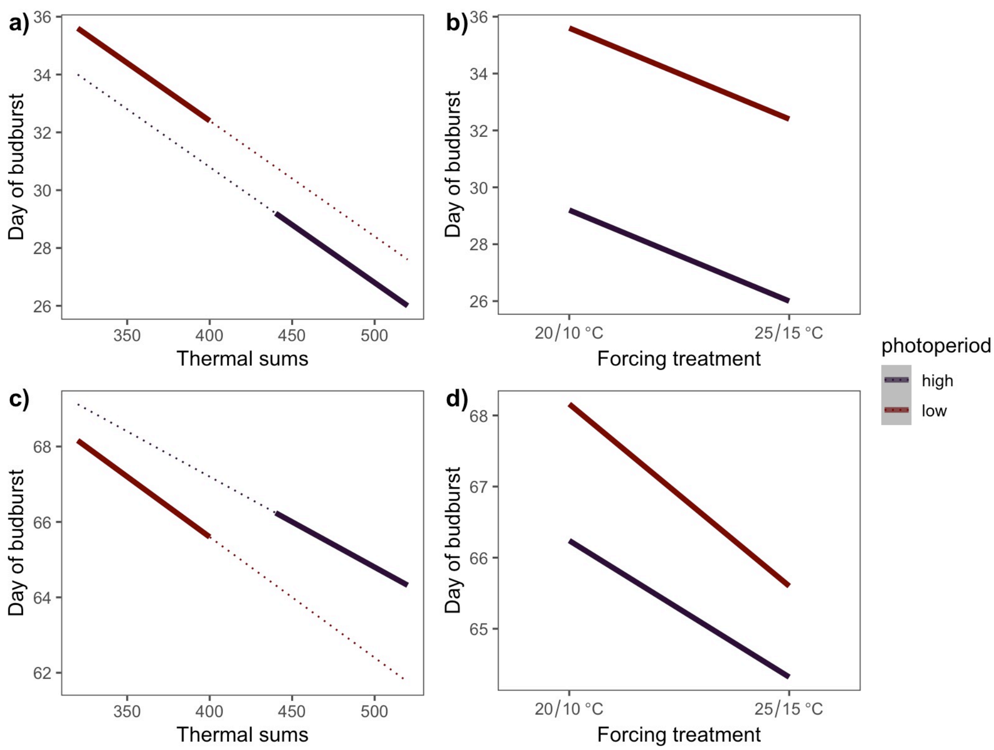

Figure 2: Estimated effects of photoperiod and forcing on spring phenology based on a simulated experiment in which the coupling of photoperiod and thermoperiod introduce an experimental covariation between the temperature and light treatments. The dotted lines in **a)** and **c)** depict the true effects of forcing at each photoperiod level, and the solid lines depict the estimated effects. **a)** depicts a scenario where forcing and photoperiod effects do not interact, while **c)** includes an interactive effect. **b)** and **d)** depict the estimated effects of forcing and photoperiod, and for **d)**, their interaction, if the experimental covariation due to periodicity coupling in **a)** and **c)** is unacknowledged, and treatments were assumed to be orthogonal.

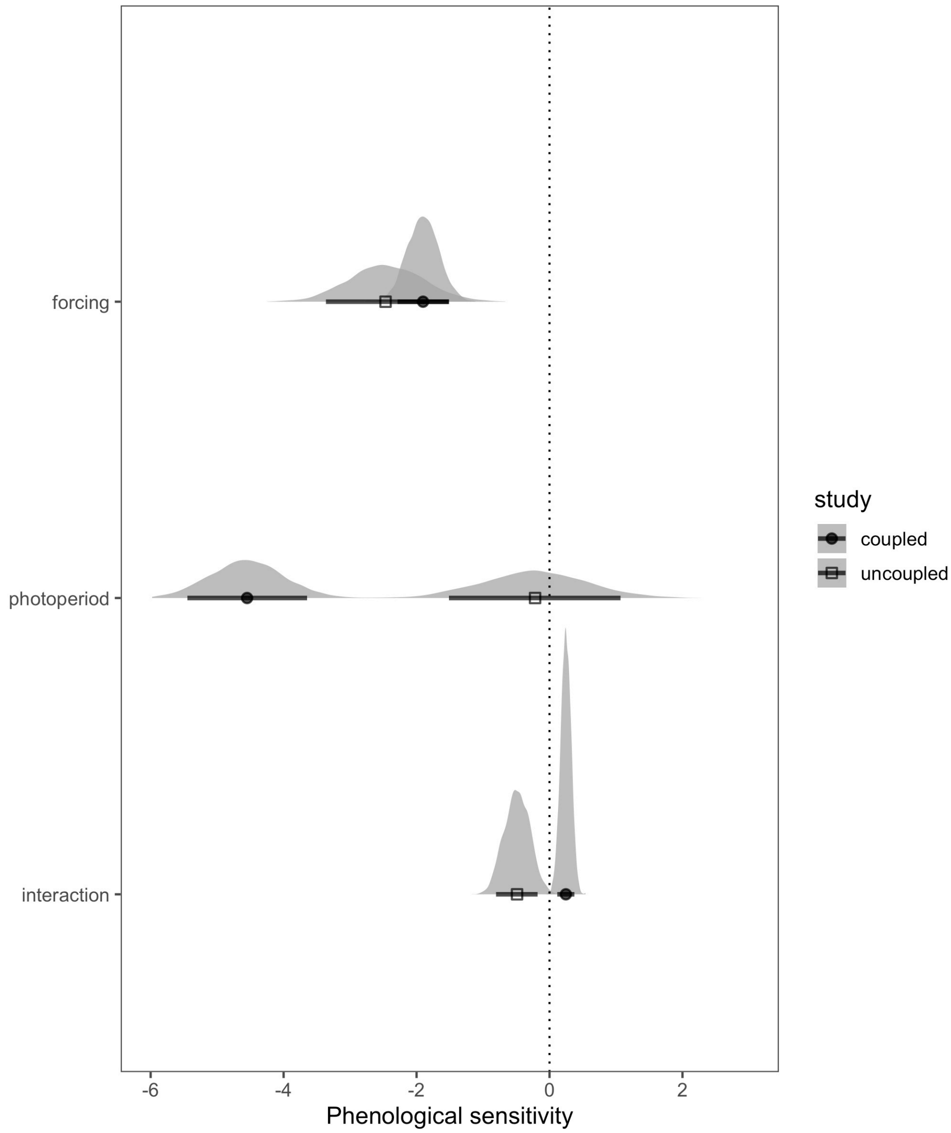

Figure 3: Estimated phenological sensitivity, (Δ day of leaf expansion/Δ unit increase in cue level) using alternative methods of varying thermoperiod relative to photoperiod. Points indicate the estimated mean effect and bars the 90% uncertainty intervals. The full posterior distributions for each parameter are also depicted as additional display of uncertainty. The coupled thermo- photoperiod design is from Flynn & Wolkovich (2018) and the uncoupled design is from Buonaiuto & Wolkovich (2021).

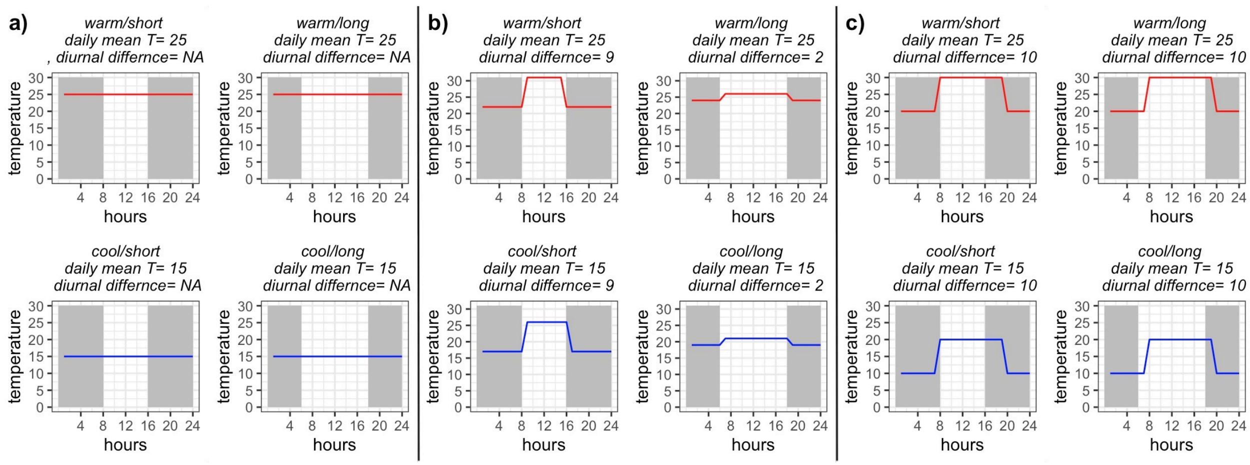

Figure 4: Conceptualized experimental designs to test temperature and daylength interactions on a biological response. Design **a)** manipulates temperature intensity only (no thermoperiodicity). In **b)**, photo- and thermo- periods are coupled, but the orthogonality of daily temperature treatments is maintained by proportionately varying the diurnal temperature fluctuations across photoperiod treatments. In design **c)**, consistent diurnal temperature fluctuations are maintained but, thermoperiod and photoperiod are decoupled and varied independently, maintaining orthogonality in daily temperature treatments.