

# Introduction

Over a century of biological research has demonstrated that across the tree of life temperature and photoperiod shape a number of important biological processes including growth and metabolic rates (), sex determination (), acclimatization to seasonal environments () and phenology (). These biological responses in turn dictate broad scale ecological processes and patterns ranging from biogeochemical cycling () to species range limits (). 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.

Because environmental factors often co-vary in the field (for example, in most temperate ecosystems photoperiod and temperature both increase as the season progresses ()) experimental manipulations of climate variables in artificial environments are ideal for mechanistically characterizing biological responses to environmental fluctuations (??). Growth chambers of all shapes and sizes have been used for over a century to separate the effects of different climate factors on biological processes () and these efforts have greatly advanced researchers ability to predict the response of organisms to current and future climate change and improved the precision of large scale climate models ().

Yet growth chamber studies come with their own unique set of challenges. When designing artificial environments, researcher must choose how many and which environmental factors to vary, the intensity and periodicity of these treatments and the background levels of other environmental factors that are not being manipulated. Because biological responses to the environment are generally the product of complex interactions between multiple environmental signals (), these seeming small choices about experimental design can generate significant differences in outcomes, and as we demonstrate below, strongly bias the subsequent analyses. Treatments appear to rarely be standardized among researchers even within disciplines () and these complexities may in part contribute to many discrepancies between experimental studies and observation data ().

Even with these limitations, growth chamber 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. Below, using plant phenology experiments as a case study, we discuss several common-use approaches to manipulating photoperiod and temperature treatments. Then using both theory and experimental evidence, we demonstrate how these subtle choices in experimental design can strongly impact modeling inferences. Finally, we provide a general framework for treatment decision making. To concretely illustrate the problems and advantages of treatment selection, we focus our discussion on the phenology of temperate woody plants, yet the insights of our case study can be readily applied to other study systems and biological processes.

## Phenology, temperature and photoperiod

Decades experimental work in growth chambers have demonstrated that temperature (winter chilling and spring forcing) and photoperiod are the primary cues of phenology for plants in the temperate/boreal zones (). Like many other biological processes, recent advances have demonstrated that plant phenological responses are nonlinear, due largely to interactions between cues(??), highlighting the need for experiments to be designed to evaluate the strength of these interactions. To truly test interactions, treatments must be orthogonal: that is there must be more than one level of each treatment, and each level of each treatment must be crossed with every level of the other. While chilling is an important cue for spring phenology(?) and interacts with other cues(?) , experiments apply chilling in complete darkness or approximate chilling by sequential field sampling dates so they are less prone to some of the complexities that arise under forcing and photoperiod treatments. Because of this we focus our discussion on experiments that manipulate photoperiod and forcing conditions, well recognizing that a good orthogonal experiment will also include chilling treatments.

For both light and forcing cues, there are two axes that can be manipulated in experiments. Then intensity of the cue (ie temperature of forcing treatment or intensity/wavelength of the light treatment) and the period

(duration of temperature exposure or light periods). While light intensity can affect phenology (??), period is considered to be the dominate light cue, and phenological research tends to focus on photoperiod rather than intensity (but see ()). For temperature, intensity is an established driver of phenology, but the very common and successful modeling framework for quantifying temperature effects on phenology, the growing degree day or growing degree hour, integrates temperature with its period (). Seemingly simple decisions about how to vary photo-period and thermo-period relative to each other can disrupt the orthogonality of an experiment and add uncertainty to experimental results. Below we present three common approaches to varying these three primary treatment axes (photo-period, thermo-period and thermo-intensity) in experiments, and discuss how each decision may affect experimental inferences.

### **Manipulation of Temperature intensity and photoperiod:**

The most simple design to test for a temperature x light interaction is to manipulate only one axis of each cue, for example, temperature intensity and light period. This would be implemented by applying high and low forcing treatments at constant temperatures, and crossing these with a long and short photo-period treatment at a constant light intensity (figure ??a).

The main advantage of this design is that it is simple to implement, maintains the orthogonality among the four treatment combinations (warm/short, warm/long, cool/short, cool/long) and allows for a very straightforward interpretation of the predictors. Additionally some studies have found that leaf expansion is faster at constant temperatures when compared to diurnally fluctuating thermoperiods (?). However, in nature, plants experience substantial diurnal temperature variation, and as such, growth chamber experiments without a diurnal thermoperiod would be a poor approximation of natural conditions. Other studies indicated that species may in fact be responding to the differences between day and night temperatures (?), which raises further questions about the utility of experiments lacking thermoperiodicity for understanding the ecology and evolution of plants in nature. On account of these shortcomings, it is more common for researchers to include a diurnally varying thermo-period for forcing treatments. For example, rather than static high/low forcing treatments of 20°C and 16°C, researchers would use, a high forcing treatment of 24°/16°C and low 20°/12°C (day/night).

### **Coupled photo- and thermo-periodicity:**

Incorporating diurnal thermo-periodicity into experimental design requires that the researcher decide how to vary this periodicity relative to the experimental photo-period cycle. The vast majority of published experiments employ a design in which photo- and thermo-periodicity are synchronized as in (figure ??b) (?). This design is perhaps the most intuitive, one, following the fact that the photo- and thermo-period tend to co-vary in nature (?). This approximation of natural conditions may make this design useful for extrapolating inferences between experimental and observational studies. However, while it appears, to maintain orthogonality between treatments when considering temperature intensity alone, but, in this case, the forcing response is a product of both the intensity of the temperature cue and the duration of its exposure (growing degree hours). Thus, the coupling of photo- and thermo-period results in non-orthogonality between the four treatment combinations (see figure ??). In this scenario, the effect of photo-period cannot be neatly distinguished from the effects of increased thermo-period under the longer day treatments, making it impossible to accurately evaluate the relative importance of these cues, which is often a major goal of growth chamber experiments. This results in spurious interpretation of experimental results.

**Decoupled photo- and thermo-periodicity:** An alternative experimental design decouples the experimental thermo- and photo- periods. This would be achieved by varying thermo-period on a constant schedule across all four treatment combinations (figure ??c). Though this experimental set up has not been widely employed in growth chamber experiments (but see (?)), this design restores full orthogonality to growing degree hour sums that make up the forcing treatment (figure ??a).

But this obvious advantage for reasonable statistical comparisons introduces a new biological artifact that may bias experimental results. Evidence from horticulture studies have demonstrated that cell growth is

most sensitive to temperature fluctuation at the beginning of the photoperiod(?). ? found that increasing temperatures in the first two hours of the photoperiod was almost as effective for stimulating shoot elongation as similar temperature increases for the whole photoperiod. More generally, recent work suggests that phenology is more responsive to daytime warming than night (?). By de-coupling thermo-period from the photo-period manipulations, experimenters by necessity introduce an asymmetry between the photoperiod treatments *relative* to the thermo-period. In our example in ??c), the long and short day photoperiod treatments have the same 12 thermoperiodicity cycle, but in the long day treatments, the plant does not experience "day time" temperatures until after they encounter "day light", while the plants in the short day treatment experience day time warming before light.

One solution to this problem is a modified decoupled design where the decoupling of thermo- and photo periodicity happens in the evening ??d) in an improved version of a uncouple thermo and photo period design, because the temporal relationship between dawn light and temperature is standardized across all treatments, but this again produces an unrealistic comparison with nature, and the "true" effects of warming may be obscured by that fact for the short day treatments, there is less overlap between day time temperature and light conditions.

As we have shown above, each design introduces its own biological or statistical artifacts that will bias the results of the experiments. We cannot conclusively solve the problems of true inference through modifying experimental designs, but rather, we must thoughtfully incorporate the implications of these design choices into our statistical metrics and interpretation of results.

While each design has a tradeoff we expect that a coupled photo-and thermo- period design will remain the most popular choice for phenological experiments in artificial environments. In the following section, we will use a public dataset to demonstrate how mathematical principles can be applied to tease apart the artifact introduced by this inherently imperfect experimental design and make more meaningful inferences about the cue interactions.

## Theoretical evidence

The following analysis is based on results from a large scale growth chamber experiment by ?, in which in addition to chilling and provenance, fully factorial light and temperature manipulations were applied to twig cuttings from 28 woody species. The authors coupled thermo-photo periodicity in the experimental setup and estimated a mean budburst sensitive to forcing of -9.5 units?, and a mean sensitivity to photo-period of -4.5 units? with a weak, negative interaction between the two cues. We set out to calculate how much of the reported photo-period response could in actuality be driven by the unacknowledged differences in thermo-period between the long and short photo-period treatments.

Our analysis rests on several assumptions. First, we assume that following the phenology literature forcing is a more dominant cue than photoperiod (). This suggests that if anything, the photoperiod cue will be over-estimated due to the differences in thermo-period as opposed to the other way around. Second because the authors only report a weak interaction between these two cues, we treat these responses as additive. Third, we assume we assume these cues are linear.

Math here

## Experimental comparison

Our prediction of "how much of the reported photo-period response could in actuality be driven by the unacknowledged differences in thermo-period" can be rephrased as a prediction of the expected difference in estimated photoperiod sensitivity between a coupled and decoupled photo- and thermo-period experiment. While we are aware of no experiments that explicitly test these different designs, we now present an additional growth chamber data set from ? that allows us to empirically test our predictions. The ? study applied

several treatments levels that overlap with those in the ? study to twig cuttings from the same source population. However in the second study the authors decoupled photo- and thermo-period. We subset each dataset to only include species and treatment levels common among them, and combined and re-analyzed them in a model that estimated differences in the photo-period effect between the two studies. We found that the estimated difference in the average response to photoperiod among study designs was on the same order our mathematical prediction.

## Moving forward

1. What to say here

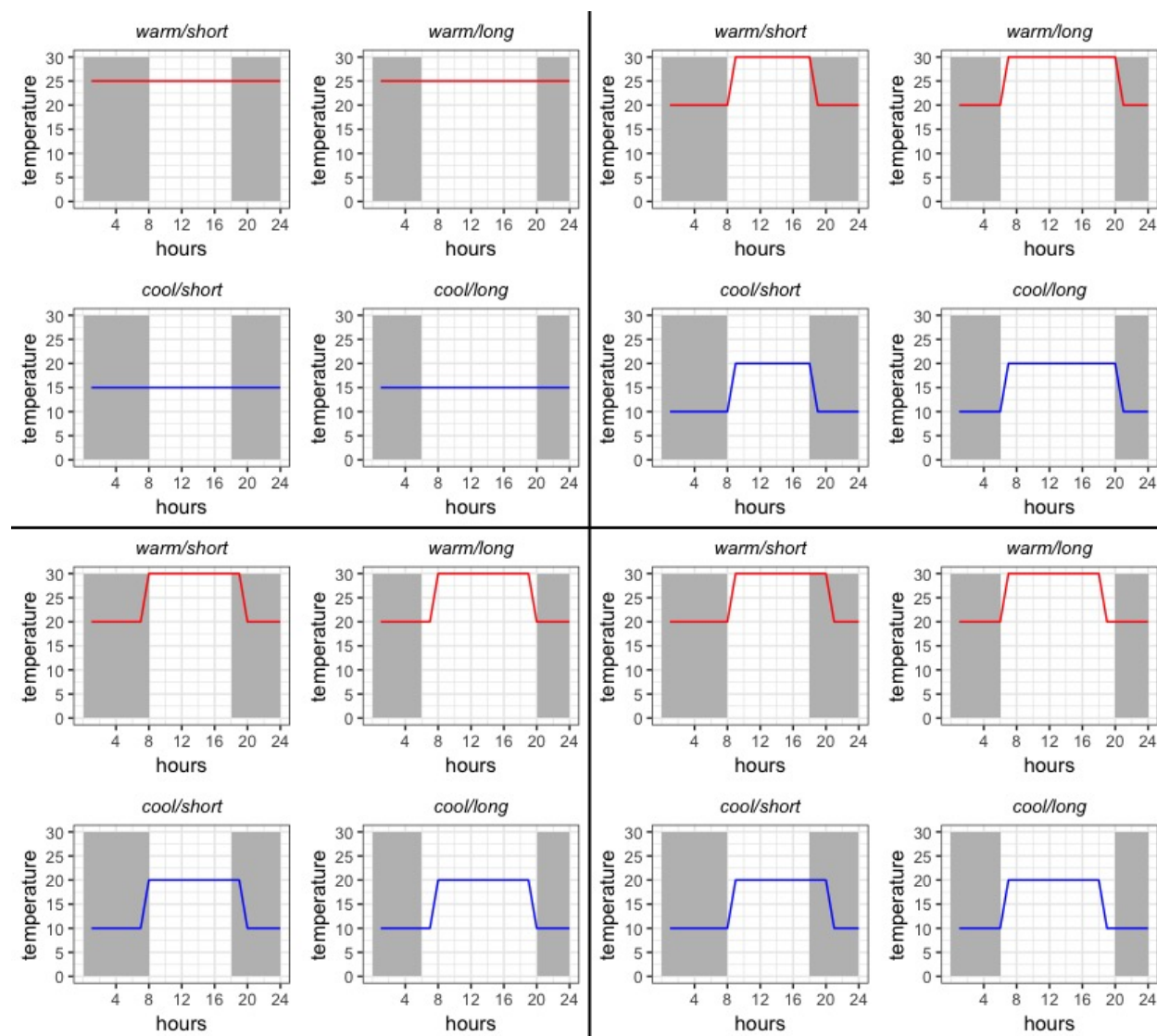


Figure 1: Experimental treatments

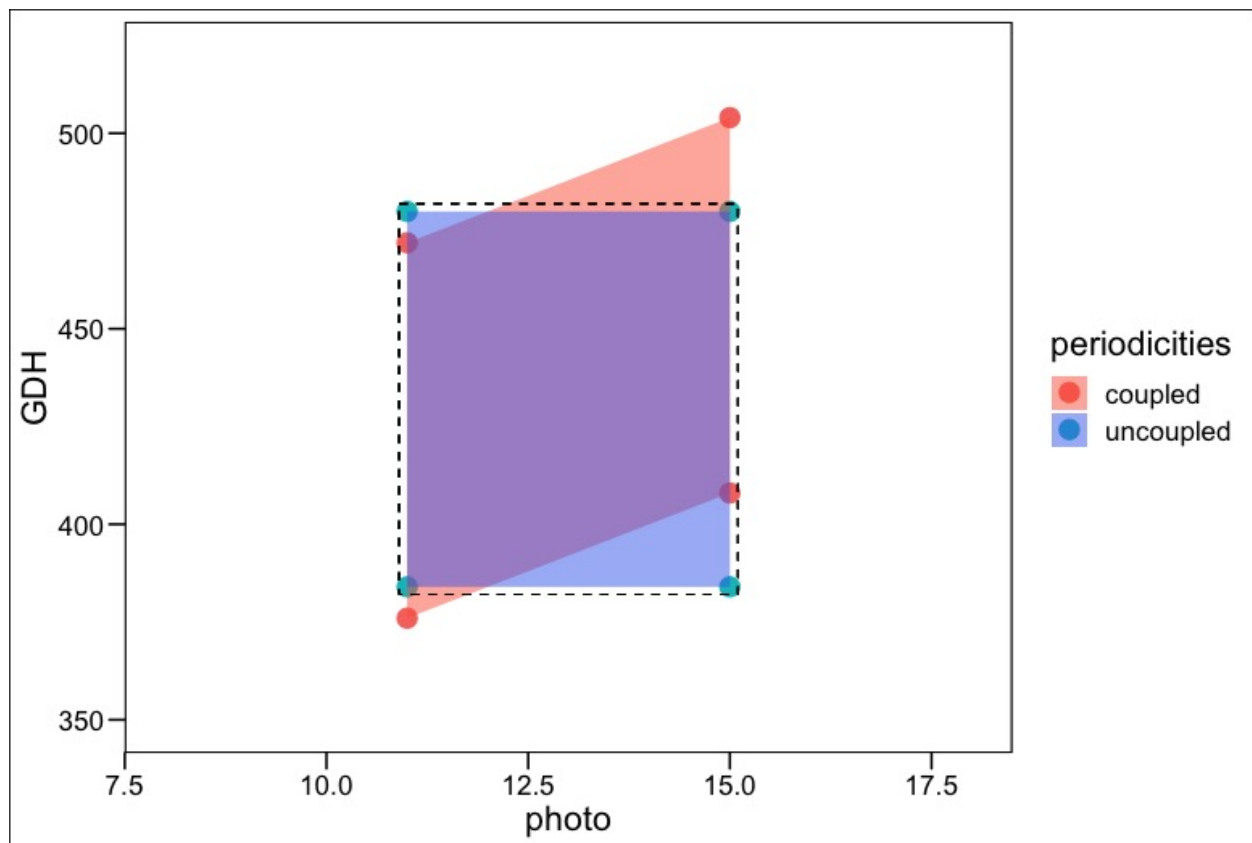


Figure 2: Experimental comparison all species

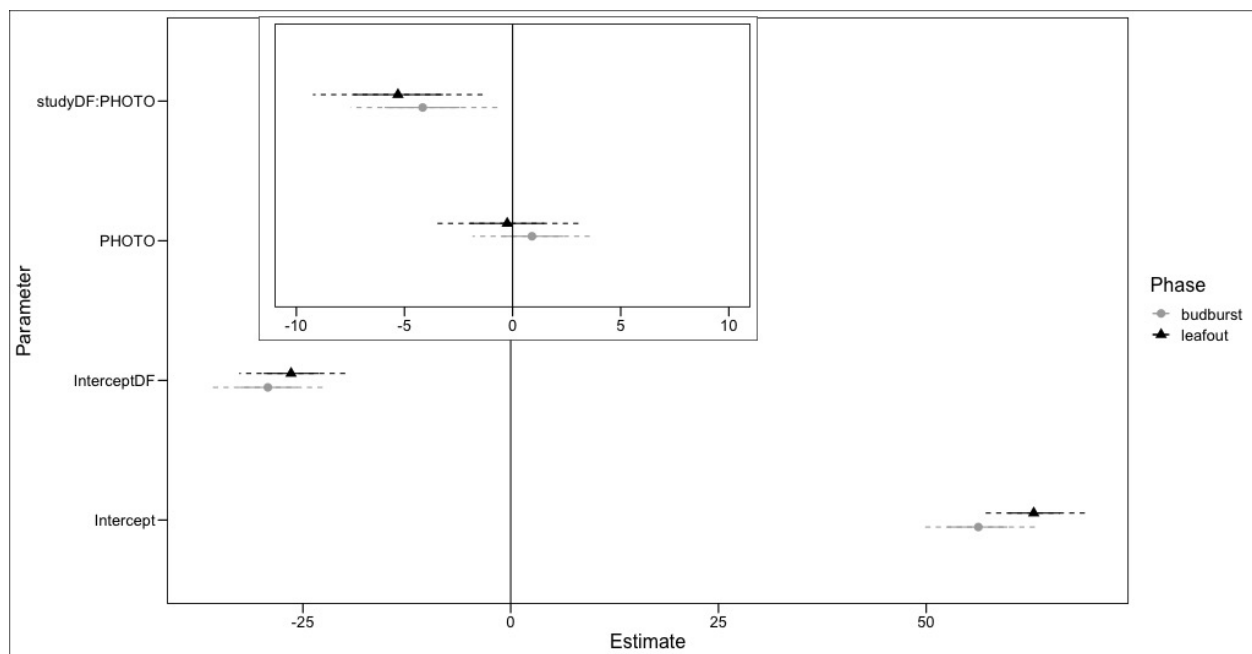


Figure 3: All species with partial pooling

	predicted difference	observed difference (sd)
bud bust	3.0	-4.1575777(2.665567)
leaf out	3.0	-5.3006250 (3.099359)

Figure 4: Table