Concept paper on understanding interactive cues and climate change (with growth chamber studies)

(1) How interactive cues will shape climate change responses

(2) Limiting cues: How spring warming, winter chilling and daylength will shape climate change responses

(3) Spring warming, winter warming or daylength: What cue will be most limiting in future tree phenology?

THE LAB[1,2]

[1] *Arnold Arboretum of Harvard University, 1300 Centre Street, Boston, Massachusetts, 02131, USA*

[2] *Organismic & Evolutionary Biology, Harvard University, 26 Oxford Street, Cambridge, Massachusetts, 02138, USA*

[3] *Forest & Conservation Sciences, Faculty of Forestry, University of British Columbia, 2424 Main Mall, Vancouver, BC V6T 1Z4*

Corresponding author: Lizzie, see [1,2] above ; E-mail:.

Abstract

Climate change has shifted plant phenology globally, with average shifts of 4-6 days° C and some species shifting several weeks. Globally, such shifts have been some of the most reported and most predictable biological impacts of climate change. This predictability comes from decades of research, which have outlined the major cues that drive most studied plant phenology: temperatures (including spring warming and winter chilling) and daylength. Further simplifying predictions, spring temperatures are often the dominant cue in nature, making linear models of heat sums often excellent at predicting interannual variation in phenology. Yet as climate change has marched on, new research has uncovered possible failures to predict the current observed changes, with many shifts appearing more muted over certain time periods or in certain locations. Here we argue that some of these inaccurate predictions are due to simple models that neglect to consider other major cues—especially winter chilling and daylength, which moderate and shape plant phenological responses to spring warming. We highlight how over 60 years of research in controlled environments can improve predictions for when, where and how the interactive effects of other cues will impact simple linear predictions. Finally, we discuss how a new generation of controlled environment experiments could rapidly improve our predictive capacity for woody plant phenology in coming decades.

Main message (and, really, it's important): If you want to project climate change impacts, you need to focus on relevant changes in all three cues. The relevant changes part is about comparing cues, the all three cues is about interactive cues.

Keywords: phenology, climate change, spring warming, forcing, chilling, daylength, photoperiod, non-linear responses, leafout, budburst

1 Main text

Shifts in spring plant phenology are one of the most reported and most predictable changes with climate change. Decades of research have documented advancing budburst, leafout and flowering across systems (Delpierre *et al.*, 2009; Yu *et al.*, 2010; Ellwood *et al.*, 2013; Jochner *et al.*, 2013; Hereford *et al.*, 2017), especially in temperate systems where long-term records have made a strong case for how humans have altered the timing of spring (Schwartz, 1997; Menzel *et al.*, 2003; **?**). Recently, however, these advances have appeared to slow (Fu *et al.*, 2015) or even reverse in some places (Yu *et al.*, 2010)—failing to match simple predictions of an advancing spring with continued warming (Ellwood *et al.*, 2013). The main hypothesis for this failure is that plants' cue to spring warming—which most observational studies focus on—is no longer the only cue that matters to predicting responses to continued warming (Chuine *et al.*, 2016).

Despite the focus of much spring phenology researach on spring warming cues the underlying physiology of phenology is much more complicated. For many species three major cues drive spring phenology: forcing (warm temperatues, generally occurring in the late winter and early spring), chilling (cool temperatures, generally occurring in the fall and winter), photoperiod (daylength). Forcing and chilling cues are generally understood to be accummulation processes, where plants must integrate chilling and forcing experienced over time to meet a sum threshold value at which they can budburst, leafout or flower. In contrast, photoperiod is generally not considered an integral cue, but one evaluated daily. These cues together may create critical non-linear responses that many current methods cannot predict. [TRANSITION?]

Measuring these three major cues and understanding how they will interactively determine current and especially future phenology is hard for several major reasons. First, these cues vary across species, and possibly within species across the range (Vitasse *et al.*, 2009; Harrington & Gould, 2015). Second, the cues appear designed to interact: one cues may compensate for another cue. In practice this means studies that are not designed to tease apart these cues experimentally (e.g., observational studies) may find statistical evidence for only one cue, because it masks one or more other cues. For example, a plant that has not received enough chilling would generally require more forcing and thus leafout later, which could look indistinguishable from a photoperiod requirement that has not been met.

To some extent the interactions of these cues have meant that—for many locations in most years—researchers have not been challenged to measure cues beyond forcing to understand advances and make accurate near-term predictions. In many locations if chilling cues are met and forcing produces budburst at an acceptably long photoperiod—then the main cue needed for predictions is simply forcing. In such cases, simple linear models of spring warming versus spring phenology may suffice. But, increasing evidence, in some systems in some years, suggests these chilling and photoperiod may play an important role in spring phenology today, with that importance growing in the future. If rising evidence suggests these cues are critical to understand current responses and predict future non-linearities, then how do we integrate them more into phenological research?

The first step towards improved phenological predictions is robust measurements of chilling, forcing and photoperiod cues. Recently, much effort has focused on estimating these cues from long-term observational data. Yet obsverational data may often fail to robustly estimate any of the three major cues due to two types of statistical issues. First, in most observational data these cues are correlated: forcing increases alongside longer days, and the response to a high chilling cue is the same predicted for a strong photoperiod cue (i.e., later spring phenology). Secondly, most observational studies focus on linear models of each cue and often without interactions between cues. This is perhaps unsurprising as the non-linearities in most plant phenological responses to most cues are expected to be linear except at extremes (and natural conditions often see only a small slice of the range of values of each cue that are possible), and given that interactions between cues are difficult to estimate—moreso when those cues are highly correlated in nature. The one method designed to robustly measure these cues is controlled environment (e.g., growth chamber) studies (Nagano *et al.*, 2012; Satake *et al.*, 2013).

Growth chamber studies of plant phenology have conducted for over 70 years and are specifically designed to understand non-linearities both in individual cues and produced by interactions between cues. In contrast to observational studies, controlled environment studies can manipulate all three cues, and extend to other cues that may important in some species or biomes (e.g., humidity, drought conditions, light spectra), and can tease out interactions between cues by experimentally decoupling them. Despite the prevalence of growth chamber studies for spring phenological cues, they have rarely been reviewed. Perhaps more suprisingly, they have been often poorly integrated into the current phenological literature on climate change. This includes in debates where they are critical—such as about the importance of photoperiod (CITES, incl Richardson).

Here we aim to help integrate the long-term literature on growth chamber studies into current phenological research on climate change more fully. We begin with a brief review how three major phenological cues for woody plant phenology will shift in coming decades with anthropogenic climate change. We then review 70 years of study on the three major phenological cues from growth chamber studies, with a particular aim to understand how much of the cue-space has been studied and how experimental treatments compare to shifts in cues caused by climate change. Based on this, we discuss how growth chamber studies can be best designed to understand these interactive cues and build more robust predictions.

Given our aim to improve understanding of current trends and forecasts we focus on early vegetative phases (budburst and leafout) of wild species, which are critical to plant growth and thus to models of carbon uptake and storage, and which have shifted most with climate change (CITES). We touch on other areas of research, which have been important to our understanding of the cues underlying phenology. In particular, research has been especially strong in model systems (e.g., *Arabidopsis, Populus*) and crops (Cesaraccio *et al.*, 2004)—with the exact phenophase of interest varying (potentially by a species' life history: more focus on germination and flowering in *Arabidopsis*, and more on leafout and budset in *Populus*). Given our focus on budburst and leafout, our review concentrates on woody species phenology, where most research has been

conducted. Most of our conclusions and suggested approaches, however, could equally apply to non-woody species and/or other phenophases with underlying similar interactive cues.

1.1 How will chilling, forcing and photoperiod shift with climate change?

Before examining how chilling, forcing and photoperiod will shift with climate change, it helps to review why species have multiple environmental cues for spring phenology—to cope with the environmental variation each species experiences across its range and across years. In the spring, selection for species to track the start of resources each year (to gain access to them before other individuals have depleted them) means that species need cues that work across the interannual variation in climate across of their ranges. This variation can come from either plasticity and/or local adaptation. Spring phenology can be fairly plastic (CITE). Especially compared to other phenophases, such as budset phenology, budburst and leafout often do not show high local adaptation across a range (CITESAItken). This means that species ideally need a set of cues that work across their range (Liepe *et al.*, 2016), though there may be variation in the importance of each cue across a range (e.g., chilling can be higher in coastal versus continental, check also Legave *et al.* (2013)), and genetic work suggests the same genes can be triggered by very different enviromental cues (e.g., vernalization cues in *Arabidopsis thaliana* CITE). Cues for species with high climatic variation across their ranges will thus need cues adapted to high climatic variation (spatially and temporally).

Temperate species that employ three cues—forcing, chilling, photoperiod—provide many diverse paths to budburst each season, especially through interactions between these cues and/or non-linearties in certain cues. Indeed, one of the most critical changes that can happen is when pre-climate change conditions generally caused budburst to occur at the extremes of some cues (e.g., high chilling and/or long photoperiods, where responses may be effectively muted, see FIGURE?) and climate change has now pushed budburst into periods where these cues are at the more linear part.

All three major cues are expected to shift with climate change (See Fig. 1), though the shifts will vary substantially across space and time. Most notably to date, warming increases the forcing plants experience each day (CITES), with more rapid shifts—and thus also greater shifts—at higher elevations and in the arctic (CITES). [Give range of warming depending on different scenarios!]. Daily minima (generally night-time temperature) will generally warm more than maxima temperatures, though this effect varies spatially (CITES). Warming across seasons is also variable (ADD more here).
Warming also translates into important shifts in chilling, though a dearth of understanding on chilling makes predicted shifts in chilling complicated. Research to date suggests chilling only accumulates in a certain range of temperatures with low (e.g., $<0°C$) temperatures generally not contributing to chilling accumulations and higher temperatures (e.g., $>12°C$) potentially decreasing previously accumulated chilling. Thus, major shifts in accumulated chilling would be predicted where temperature regimes that were previously too low to accumulate chilling in many months of the winter warm such that chilling now accumulates in those months (CITES).

Areas with this shift would then expect much earlier budburst with warming, potentially far earlier than last frost dates (CITES). Unfortunately, these predictions are based on models developed almost solely for agricultural crops (but see HARRINGTON CITES etc.), especially stone fruits, and have rarely been robustly adapted to forest trees. While the development of classic models of chilling for peaches and related fruit trees benefited from data on these species being planted far outside their range into regions with extremely low or potentially no chilling, equivalent data on forest trees is almost never available. Thus chilling models to date generally use limited observational and experimental data from forest trees to try to reparametrize the basic stone fruit models (CITES).

Improved chilling models for forest trees (and horticultural trees) may come from an improved understanding of dormancy at the physiological level. Chilling is defined as what leads to break of endodormancy (DEFINE/check this definition), after which plants enter ecodormancy when accumulated forcing then leads to budburst. Measuring endodormancy and its transition into ecodormancy is notoriously difficulty, however, and work to date has relied generally on cellular staining methods tested on a very limited subset of species (CITES). These methods themselves rely on a still-developing understanding of what causes dormancy at the cellular level (CITES). In practice, most phenology studies use the terms 'chilling' and 'forcing' to mean 'cool temperatures' (often either in the fall and winter or applied first in experimental conditions) and 'warm temperatures' (often either in the spring or applied after sufficent chilling) and generally hope they correspond to endo- and eco-dormancy. Some studies use the sequential transfer of cuttings to warm conditions to estimate the transition from endo- to eco-dormancy, with rapid and full (e.g., >90% of buds on a cutting) budburst generally meaning a plant is ecodormant, but given that this is labor- and space-intensive most studies considering chilling do not do this. We expect that without a much improved physiological understanding of endo-dormancy release robust models of chilling—and thus accurate predictions of how chilling will shift with climate change—will be difficult to develop.

Research has discussed shifts in photoperiod with climate change far less often compared to predicted changes in forcing and chilling (but see CITES). While an environment's photoperiod does not shift with climate change, the relevant photoperiod a plant experiences at critical physiological points may change dramatically with warming. In particular, increases in chilling and/or forcing, which could alone produce much earlier budburst, may be offset by short photoperiods (CITES) that delay budburst. Similarly, long photoperiods can lead to budburst sooner than predicted by solely low chilling or forcing conditions (CITES). Thus, changes in chilling and/or forcing correspond to changes in the relevant photoperiod with climate change.

These cues may create non-linear responses with warming in three major ways. First, as discussed above, each cue may be non-linear alone. For example, cues may be linear in the mid-range of values, while extremely high or low values of some cues (e.g., photoperiod) may produce alternative response. For example, at very low photoperiod (short days) plants often will not grow, similarly maximum growth may occur at photoperiods of 18 hours, meaning photoperiods longer than 18 hours will have no additional effect on budburst timing. Secondly, interactions

6

between cues can produce non-linearities. For example, if forcing and photoperiod both determine budburst timing and have a subadditive interaction (i.e., both cues together produce a more muted response than the addition of each cue's effect alone) then the two cues together will produce a non-linear response. The interaction of interactive cues with common environmental correlations may cause additional non-linear complexities (e.g., warm springs may cause mean plants meeting their forcing cues at shorter photoperiods).

Given our current understanding of phenological cues, we should expect the most dramatic changes in phenology in systems with non-linear responses where climate change pushes the system across a critical threshold or inflection points. For example, if some species have a critical photoperiod for budburst and warming means forcing cues are met before the critical threshold, then we would expect incomplete or highly delayed budburst (CITES). Similarly the complexity of chilling could produce myriad non-linearities. If, as currently hypothesized, chilling follows an optimum curve (i.e., chilling does not accumulate at very low or high temperatures but in between it accumulates at a maximum rate at some temperature optimum), then endodormancy break would shift earlier in systems where warming pushes winter temperatures into a chilling-accumulation temperature or closer to the optimum temperature, and delay where warming pushes winter temperatures beyond the optimum (with more complex predictions if high temperatures decrease accumulated chilling). Interactive cues could also produce non-linear responses, but predicting these requires a refined understanding of the interaction and whether there are critical inflection points that may crossed with continued warming. These complexities highlight how difficult predictions may be without careful efforts to tease out how each cue works alone and interactively.

1.2 Predicting non-linear responses with controlled environment experiments

Controlled environment (generally growth chamber students) can help predict non-linear responses by allowing researchers to examine the effects of one cue with the others held constant, and examine interactive effects (given the appropriate study design). Such experiments may be especially useful for forecasting if they are designed in a range relevant to current versus future conditions. Indeed, one of the major advantages of experiments is that they allow treatments outside of the historical range—an option long-term observational data can never provide. We reviewed controlled environment studies over the last seven decades (1947-2014) to understand the range of treatments already available, and how they compare to current and future conditions. We note that these studies were rarely conducted for climate change research, and most often done for fundmental science or other areas of applied science (e.g., horticulture or forestry). Nevertheless, they are some of the best available data for how plants respond to the environment and thus a critical resource for climate change research today.

Controlled environment studies have been conducted across 227 species across the globe, with the majority of papers report research occurring in Europe (54 of 91 papers), followed by North America (23, Fig. 2). Most studies manipulate one cue (Fig. 3), though studies of two or three cues have occurred in almost every decade. Forcing and photoperiod were the most commonly studied cues (56% manipulated forcing; 55% manipulated photperiod), with chilling being stud-

ied in only a third of all studies (33% manipulated chilling). The actual cues studied varied across latitude with a general trend toward examining more extreme values at higher latitudes. Thus, forcing and chilling treatments (° C) declines 0.1°C per ° latitude (for forcing, min is -0.12, for max it's -0.08, see Fig 4; for chilling it's -0.1 for min and -0.07 for max); and the maximum studied photoperiod increases with latitude (0.08 hr per degree ° latitude). These shifts across space make sense—higher latitudes experience colder temperatures and longer photoperiods— but introduce a bias in results as any comparisons of studies from lower and higher latitudes are also comparing a different range of cues generally.

Single cue studies were the most common, but prevent understanding interactions among cues or comparisons of which cues dominate phenological responses; studies of multiple cues can overcome these challenges. [NEED TO CHECK] Of the studies manipulating at least one cue, half additionally manipulated another cue. Study designs most often allowed examining whether cues were interactive (that is, whether the effect of one cue depends on the level of the other cue), with the most studies testing for interactions between photoperiod and forcing (14 studies), followed by studies that examined the effects of photoperiod (12) and forcing (11) across the fall-winter. Such studies follow the design generally attributed to Weinberger (CITE) where tissue (e.g., cuttings) are taken progressively across the fall and/or winter seasons then exposed to controlled environment conditions. These studies often equate tissue removed later from the field as having received more chilling and thus often treat 'time of cutting' as interchangeable with 'chilling,' though forcing and photoperiod conditions also change. Studies examing photoperiod or forcing crossed with experimental chilling treatments (either through changes in temperature or days of chilling) were slightly less common (8 studies each for photoperiod and forcing). Studies examining three cues directly were very rare: we identified only two studies examining all three cues at once, and both were on *Picea abies* (CITE). A slightly larger set of studies (five) examined three cues indirectly—manipulating photoperiod and forcing in controlled environments but equating chilling with sequential removal of tissue from the field—for 11 species (CITES).

The paucity of studies examining multiple cues limits our fundamental understanding of each cue, as well as how—when combined—they will determine future leafout with continued warming. Because the cues are all known to be interactive, estimates of any one cue are influenced by the level of each other cue. Knowing the level of each other cue is difficult both because they are often not reported, and also because they are somewhat impossible to know given our current understanding of endodormancy (when we understand chilling is applied) and ecodormancy (when we understand forcing is applied). Authors may use the terms 'chilling' and 'forcing' for their treatments, but they rarely have physiological evidence that these are the actual treatments being applied. Studies using sequential removal from the field to estimate chilling are at perhaps the greatest disadvantage to estimate the cues applied: chilling must rely on field estimates from models that are currently only hypotheses of actual chilling (CITES), and forcing and photoperiod treatments are most probably a mix conditions in the field and conditions applied in controlled environments. Further any study conducted over much of the fall and winter is most likely estimating the effect of their 'forcing' and photoperiod treatments on first

endodormant plants, and then ecodormant plants—with when the switch happened generally unknown. Some studies do attempt to assess dormancy state ... [we need to add more here!]. [... end on needing more studies of interactive cues or just more studies reporting dormancy state and best guess at other cues?]

The utility of controlled environment studies to forecasting also depends on how relevant treatments are to current and future conditions. Estimating such relevance is difficult as it depends on a species' range and projections considered. However, a, simple analysis of two widely studied species, *Fagus sylvatica* and *Betula pendula* suggests experiments have generally bracketed the range of projected temperatures. Projected changes in maximum temperatures generally fit within the range of temperature differences conducted within forcing treatments in experiments (e.g., an experiment with both 16°C and 20°C forcing treatments would have a 4°C difference), and similarly matched differences in minimum temperatures in chilling treatments. As we noted above, however, there is a paucity of chilling studies that directly manipulate temperature—and thus allow a comparison of how differences in chilling temperatures impact phenology. Indeed, we found no studies with multiple chill temperatures tested for *Fagus sylvatica*, even though it is one of the most well-studied species.

Experimental shifts were generally larger than expected shifts due to climate change. This makes sense from an experimental-statistical perspective: if the goal of an experiment is to identify if a cue is present then larger treatment differences should yield larger effect sizes and thus higher statistical power. But such large shifts may be risky to extrapolate from into smaller shifts due to warming. Further, experimental studies vary from natural settings in myriad ways. Different studies have ameriolated some of these differences. For example, most studies (x=XX) have constant day/night temperatures, but some vary day and night temperatures (X=XX) with nights generally being cooler, while some have even introduced ramped temperature through the day and across an experiment's length (CITES). Yet these shifts are generally introduced in experiments that are not designed to test the effect of a controlled environment more similar to natural conditions (but see CITES) and thus provide little insight on how much such experimental artifacts matter. Such differences mean extrapolating from controlled environment studies should be done with care, and future studies can critically design experiments better able to improve forecasting effects of climate change.

1.3   Paths forward

We argue that controlled environment experiments will be critical for accurate predictions of phenology given future warming, How accurate such predictions are will depend on the design of future experiments and how well they can can be integrated with long-term data to improve models.

Improving controlled environment studies

We expect the most useful future experiments will be designed to improve models. In particular, experiments designed to identify threshold effects and optimal temperatures/photoperiods, and non-linearities from interactive cues may be most useful. Identifying threshold effects and optimal temperatures or photoperiods generally requires many different levels of a single cue, which

can make such experiments difficult to cross with other cues. Yet, understanding if findings are consistent across varying levels of other cues should be a follow-up step to confirm that findings can be applied across levels of other cues. Studies manipulating more than one cue also test for non-linearities due to interactive cues, as long as they are properly designed: fully-crossed design (i.e., every combination of levels is present in treatments; for example, a study with two photoperiod and three forcing levels would need six different treatments). Such experiments can quickly require a large number of controlled environments, but provide critical information for models and to connect to long-term findings. As growing experimental results support that all cues are dependent on the level of other cues (CITES) and long-term data hint at multiple cues, we believe this should be a major research aim.

Controlled environment studies may also be more relevant to forecasting by exploring more realistic conditions. While identifying thresholds, optima and non-linearities may involve considering informative extremes in levels of cues, most changes in cues due to climate change are on a (relatively) smaller scale (Fig XX, CITES). Thus, experimentalists should examine cues within the current and projected future range of species when designing studies to contribute to improved forecasting of particular species. In most species distribution models, species are expected to remain in the same climatic conditions, suggesting there may be minimal changes—assuming such models of species distibution are accurate and that species track perfectly. To date, most evidence suggests species will lag in their spatial responses (CITES), meaning shifts in cues in the current range may be important to the fate of trailing edge populations (CITES). Beyond the absolute level of cues, controlled environment studies need more work on what attributes of the design are more or less critical for replicating responses from the field. For example, studies have shown differing day/night temperatures are important for some species (CITES), but comparison studies have not been conducted for most species. Equally, a few studies have attempted to replicate certain aspects of the environment, such as fluctuating temperatures, ramped temperatures and light throughout the day, the coinicidence of temperature and sunrise (CITES), but these are by no means widespread enough to understand how important these conditions are for extrapolation to models.

Improving integration of controlled environment studies with long-term data

A major need currently is improved integration between long-term data and controlled environment studies. With important exceptions (CITES), studies of long-term observational phenology data have moved forward independently from controlled environment studies. Recent observations of declining responses to warming in long-term data suggest cues beyond forcing may be playing a role (CITES) highlight the major need for non-linear models based on interactive cues. To this aim, a number of recent papers have attempted to identify chilling or photoperiod cues in long-term data. Correlations between these predictors, however, are almost always too high for any useful analysis (CITES). Other approaches that attempt to break some of these correlations, such as leveraging elevational or latitudinal gradients, may run afoul of other correlations—for example, gradients in local adaptation that mean cues shifts along these gradients in complex ways (CITEPhillimore). Similarly, controlled environment studies, as we have reviewed here, generally do not use long-term data to help interpret results or define treatments.

While most studies of long-term data and controlled environments generally ignore one another, attempts to integrate the two provide an important path forward (Caffarra *et al.*, 2011; Nagano

et al., 2012; Satake *et al.*, 2013; Ford *et al.*, 2016; Chuine & Regniere, 2017). Experiments that test for thresholds and the presence of important interactions have helped re-design models (Caffarra *et al.*, 2011; Chuine & Regniere, 2017), while other studies have used experiments to test extremes (e.g., extremely low chilling) combined with data from long-term provenance studies to understand how growth and phenology will combine to determine future ranges (Ford *et al.*, 2016). Further, some work has used controlled environments to test model predictions, especially in future climate scenarios where non-linearities are predicted (see Nagano *et al.*, 2012). Such work underlies progress towards model development that relies continuously on a back-and-forth process between developing models based on both long-term data and experiments, then testing predictions with new experiments and as newer observational data are generated (i.e., more years and also data from new locations) (Nagano *et al.*, 2012; Satake *et al.*, 2013). Such efforts of continual development take extensive data and thus have only been developed for a very few species (XX CHECKpapers, am I correct?).

Building species-rich predictions

Giving the efforts and data involved in models for a single species, building up to multi-species predictions may appear daunting, but multi-species models are critical for accurate forecasts that can apply to regions and large-scale vegetation models. Addressing this issue requires, of course, more data. Long-term data is generally more species-rich than controlled environment studies: compare PEP and NECTAR records which together have multi-site data on XX species, while our review of controlled environment studies found more than one study on only XX species. Thus, more diverse controlled environment studies may be the current major data limitation. Beyond data, however, new modeling approaches can help integrate current and future data more powerfully.

Bayesian hierarchical models are specifically designed for analysis of diverse datasets. With the right information and sufficient data, they can help attribute variation across studies to the species studied, the cues (i.e., chilling, forcing and photoperiod levels studies) and remaining unmeasured variation in studies (i.e., differences in chamber design may be captured by a 'study ID' variable in such models). Such models are extemely powerful for building species-rich predictions as they leverage data across all species into one model designed to capture both the cross-species and cross-study overall effects as well as species-level differences. Yet, like all models, they are more robust with more data. In particular, attributing variation due to study versus species requires the same species to be studied across several studies, which is currently not the case for most species, according to our literature review (cite Table). Thus, these models will be most useful given greater efforts to publish data (CITES). Given proper data reporting (i.e., all cue conditions must be defined, even when not manipulated, and controlled environment conditions should be fully described including relative humidity and irradiance) all studies—whether designed to improve models or forecasting, or not—can be included in such models.

1.4   Right now: It's your tomorrow

Phenology research had been conducted for centuries before anthopogenic climate change caused earlier budburst and leafout across much of the globe. Decades of controlled environment studies contributed to our fundamental understanding of the drivers of spring plant phenology today. Today, climate change requires leveraging these decades and centuries of research for more accu-

rate predictions that can help humans adapt to warming. While ultimately, thermal tolerances and ecological interactions may provide ultimate filters on which species survive warming, but phenology has proved an important and informative filter (CITES).

If researchers could better harnass the power of controlled environment experiments it could transform our fundamental understanding of this important aspect of plant biology and advance forecasting. Controlled environment studies can critically rule out (or support) hypotheses to explain observed discrepancies in long-term data and open up new pathways to use long-term data to understand current trends, helping the field move beyond trying to tease out cues using only long-term data where cues are inherently correlated. While understanding, modelling and predicting interactions among cues and their effects on phenology is challenging, any advances would yield more accurate predictions, with valuable implications to more realistically assess the effects of climate change on plant biodiversity, including agricultural and forest species. *Conclusions*

2 References

Caffarra, A., Donnelly, A. & Chuine, I. (2011) Modelling the timing of *Betula pubescens* budburst. II. Integrating complex effects of photoperiod into process-based models. *Climate Research* **46**, 159–170.

Cesaraccio, C., Spano, D., Snyder, R.L. & Duce, P. (2004) Chilling and forcing model to predict bud-burst of crop and forest species. *Agricultural and Forest Meteorology* **126**, 1–13.

Chuine, I., Bonhomme, M., Legave, J.M., García de Cortázar-Atauri, I., Charrier, G., Lacointe, A. & Améglio, T. (2016) Can phenological models predict tree phenology accurately in the future? The unrevealed hurdle of endodormancy break. *Global Change Biology* **22**, 3444–3460.

Chuine, I. & Regniere, J. (2017) Process-based models of phenology for plants and animals. *Annual Review of Ecology, Evolution, and Systematics* **48**, 159–182.

Delpierre, N., Dufrene, E., Soudani, K., Ulrich, E., Cecchini, S., Boe, J. & Francois, C. (2009) Modelling interannual and spatial variability of leaf senescence for three deciduous tree species in France. *Agricultural and Forest Meteorology* **149**, 938–948.

Ellwood, E.R., Temple, S.A., Primack, R.B., Bradley, N.L. & Davis, C.C. (2013) Record-breaking early flowering in the eastern united states. *Plos One* **8**, ellwood, Elizabeth R. Temple, Stanley A. Primack, Richard B. Bradley, Nina L. Davis, Charles C.

Ford, K.R., Harrington, C.A., Bansal, S., Gould, Peter, J. & St. Clair, J.B. (2016) Will changes in phenology track climate change? A study of growth initiation timing in coast Douglas–fir. *Global Change Biology* **22**, 3712–3723.

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.

Harrington, C.A. & Gould, P.J. (2015) Tradeoffs between chilling and forcing in satisfying dormancy requirements for pacific northwest tree species. *Frontiers in Plant Science* **6**, 120.

Hereford, J., Schmitt, J. & Ackerly, D.D. (2017) The seasonal climate niche predicts phenology and distribution of an ephemeral annual plant, mollugo verticillata. *Journal of Ecology* **105**, 1323–1334.

Jochner, S., Caffarra, A. & Menzel, A. (2013) Can spatial data substitute temporal data in phenological modelling? a survey using birch flowering. *Tree Physiology* **33**, 1256–1268.

Legave, J.M., Blanke, M., Christen, D., Giovannini, D., Mathieu, V. & Oger, R. (2013) A comprehensive overview of the spatial and temporal variability of apple bud dormancy release and blooming phenology in western europe. *International Journal of Biometeorology* **57**, 317–331.

Liepe, K.J., Hamann, A., Smets, P., Fitzpatrick, C.R. & Aitken, S.N. (2016) Adaptation of lodgepole pine and interior spruce to climate: implications for reforestation in a warming world. *Evolutionary Applications* **9**, 409–419.

Menzel, A., Jakobi, G., Ahas, R., Scheifinger, H. & Estrella, N. (2003) Variations of the climatological growing season (1951–2000) in Germany compared with other countries. *International Journal of Climatology* **23**, 793–812.

Nagano, A.J., Sato, Y., Mihara, M., Antonio, B.A., Motoyama, R., Itoh, H., Nagamura, Y. & Izawa, T. (2012) Deciphering and prediction of transcriptome dynamics under fluctuating field conditions. *Cell* **151**, 1358–1369.

Satake, A., Kawagoe, T., Saburi, Y., Chiba, Y., Sakurai, G. & Kudoh, H. (2013) Forecasting flowering phenology under climate warming by modelling the regulatory dynamics of flowering-time genes. *Nature Communications* **4**.

Schwartz, M.D. (1997) Spring index models: An approach to connecting satellite and surface phenology. *Phenology in Seasonal Climates I* (eds. H. Lieth & M.D. Schwartz), pp. 22–38, Backhuys Publisher, Leiden, The Netherlands.

Vitasse, Y., Delzon, S., Bresson, C.C., Michalet, R. & Kremer, A. (2009) Altitudinal differentiation in growth and phenology among populations of temperate-zone tree species growing in a common garden. *Canadian Journal of Forest Research-Revue Canadienne De Recherche Forestiere* **39**, 1259–1269.

Yu, H.Y., Luedeling, E. & Xu, J.C. (2010) Winter and spring warming result in delayed spring phenology on the Tibetan Plateau. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 22151–22156.
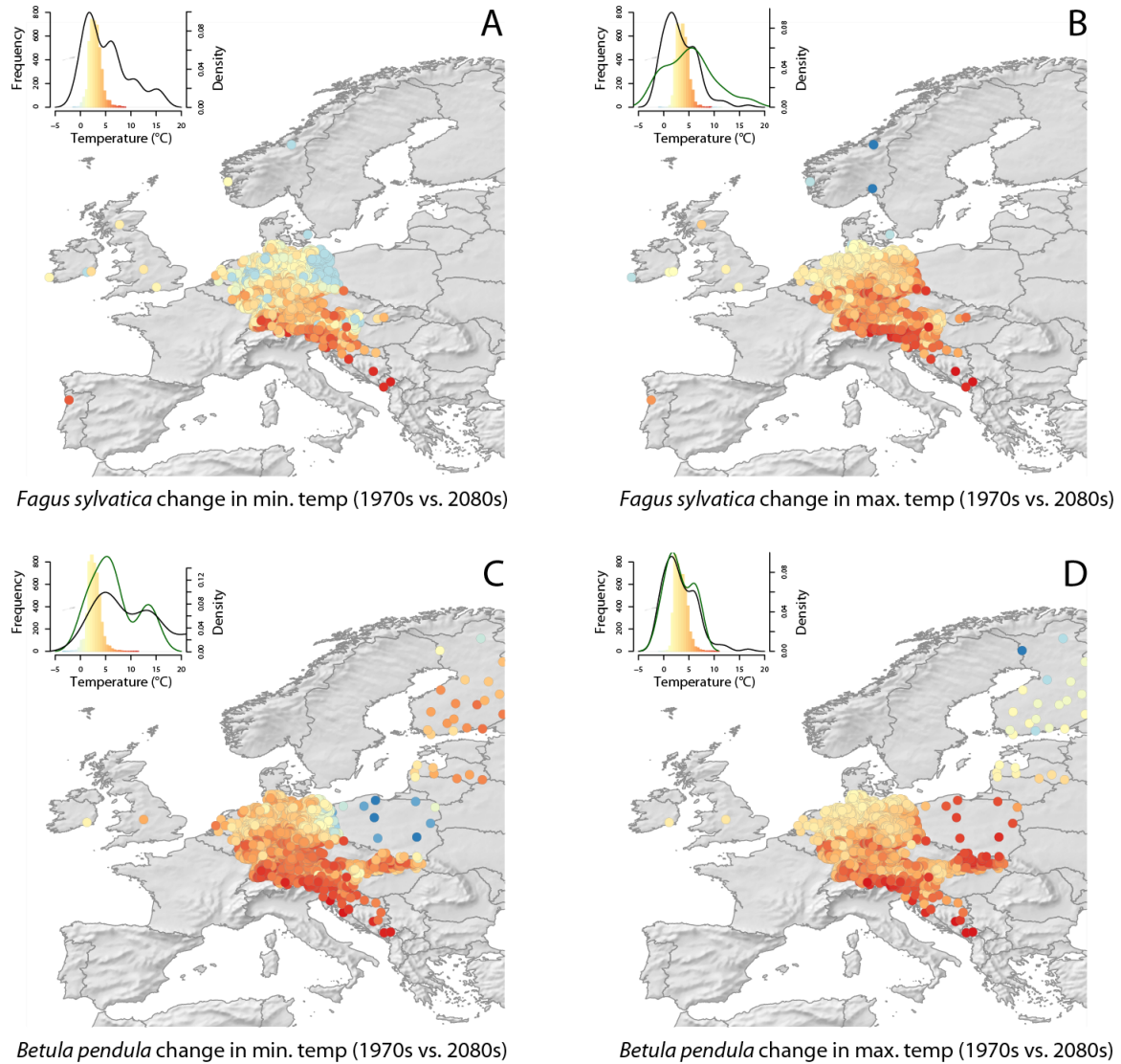
3   Figures

Figure 1: Predicted changes in temperatures relevant to chilling (A, C) and forcing (B, D) compared to a 1970s baseline shown for two species: *Fagus sylvatica* (A-B) and *Betula pendula*. Points represent a PEP725 site with XX data. Inlay plots in the upper left-hand corner of each plot show a histogram of the predicted changes in temperature overlaid with densities of the chilling (A, C) and forcing (B, D) treatments (green lines show the treatments for that exact species, while black lines show across all species; note that for *Fagus sylvatica* there are no chilling treatments of differing temperatures).

Number of Species
- 1-5
- 6-10
- 11-20
- 21-30
- 31-100
- >100
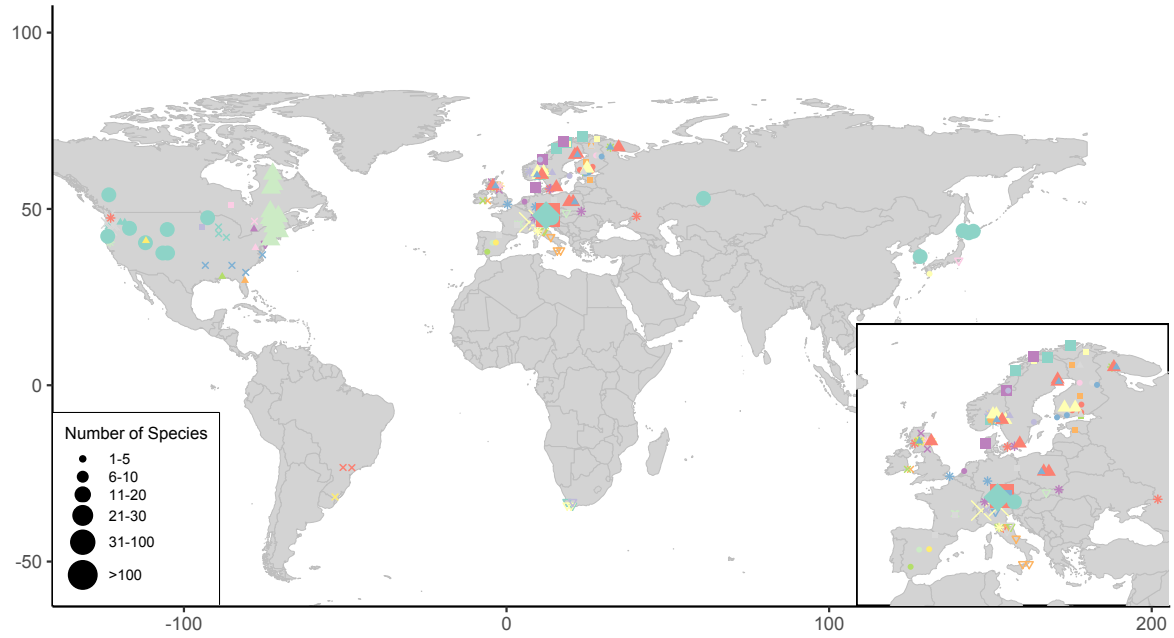
DatasetID

|			
ashby62 ; 3	gansert02 ; 1	lamb37 ; 1	ruesink98 ; 1
basler12 ; 28	ghelardini10 ; 6	laube14a ; 36	Sanz-Perez09 ; 2
basler14 ; 4	gianfagna85 ; 1	laube14b ; 9	sanzperez10 ; 3
biasi12 ; 2	gomory15 ; 2	li05 ; 3	schnabel87 ; 2
boyer ; 4	granhus09 ; 1	linkosalo06 ; 1	skuterud94 ; 6
caffarra11a ; 4	guak98 ; 1	man10 ; 1	sogaard08 ; 1
caffarra11b ; 1	guerriero90 ; 1	morin10 ; 3	sonsteby13 ; 8
calme94 ; 3	Heide03 ; 3	myking95 ; 6	sonsteby14 ; 5
campbell75 ; 6	heide05 ; 4	myking97 ; 1	spann04 ; 1
cannell83 ; 2	heide08 ; 5	myking98 ; 3	spiers74 ; 1
charrier11 ; 3	heide11 ; 3	nienstaedt66 ; 14	swartz81 ; 1
chavarria09 ; 1	heide12 ; 3	nishimoto95 ; 3	viheraaarnio06 ; 2
cook00b ; 2	heide15 ; 1	pagter15 ; 2	webb78 ; 3
cook05 ; 2	heide93 ; 9	partanen01 ; 3	worrall67 ; 12
cronje03 ; 1	heide93a ; 5	partanen98 ; 3	yazdaniha64 ; 1
falusi03 ; 1	howe95 ; 2	pettersen71 ; 1	zohner16 ; 144
falusi90 ; 1	jones12 ; 3	ramos99 ; 1	
falusi96 ; 5	junttila12 ; 6	rinne94 ; 1	
falusi97 ; 2	karlsson03 ; 2	rinne97 ; 1	

Figure 2: Overview of the data across space.
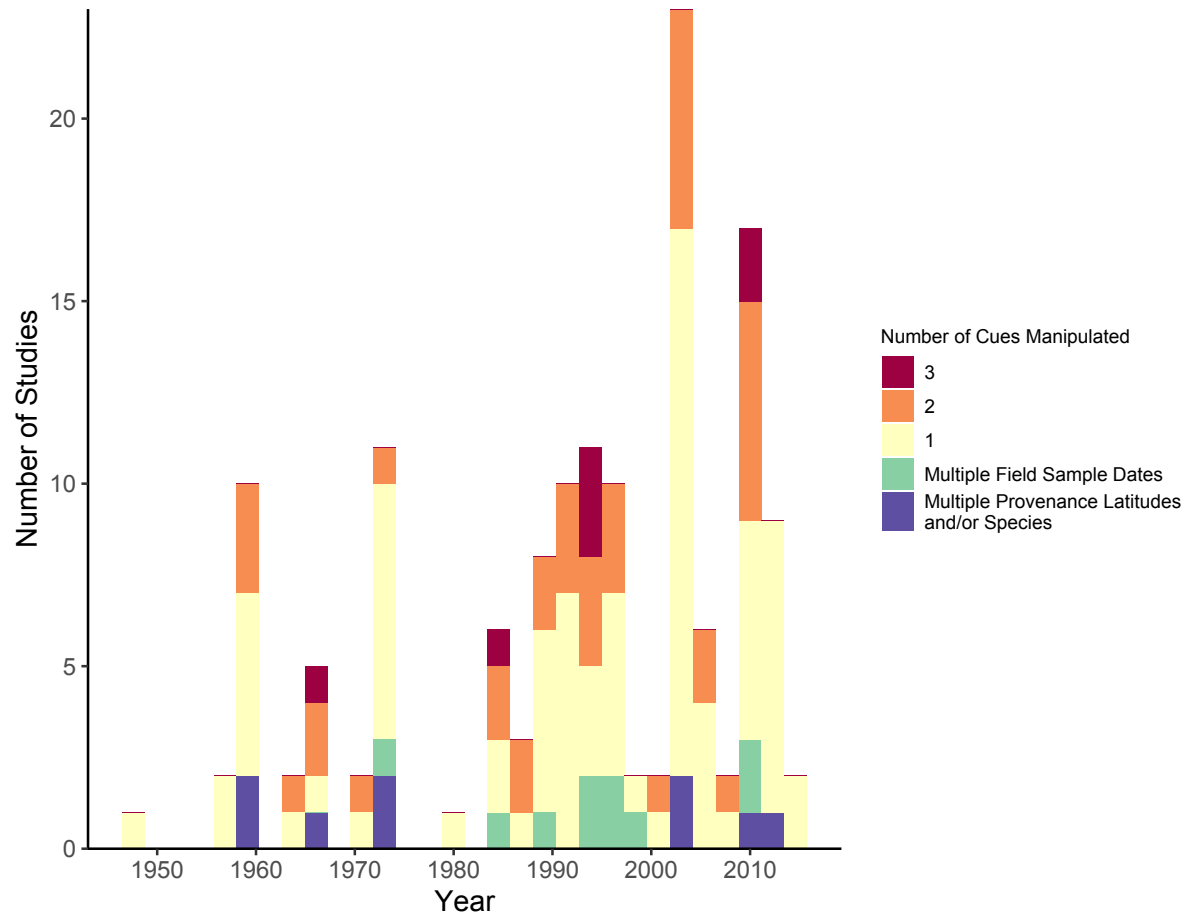
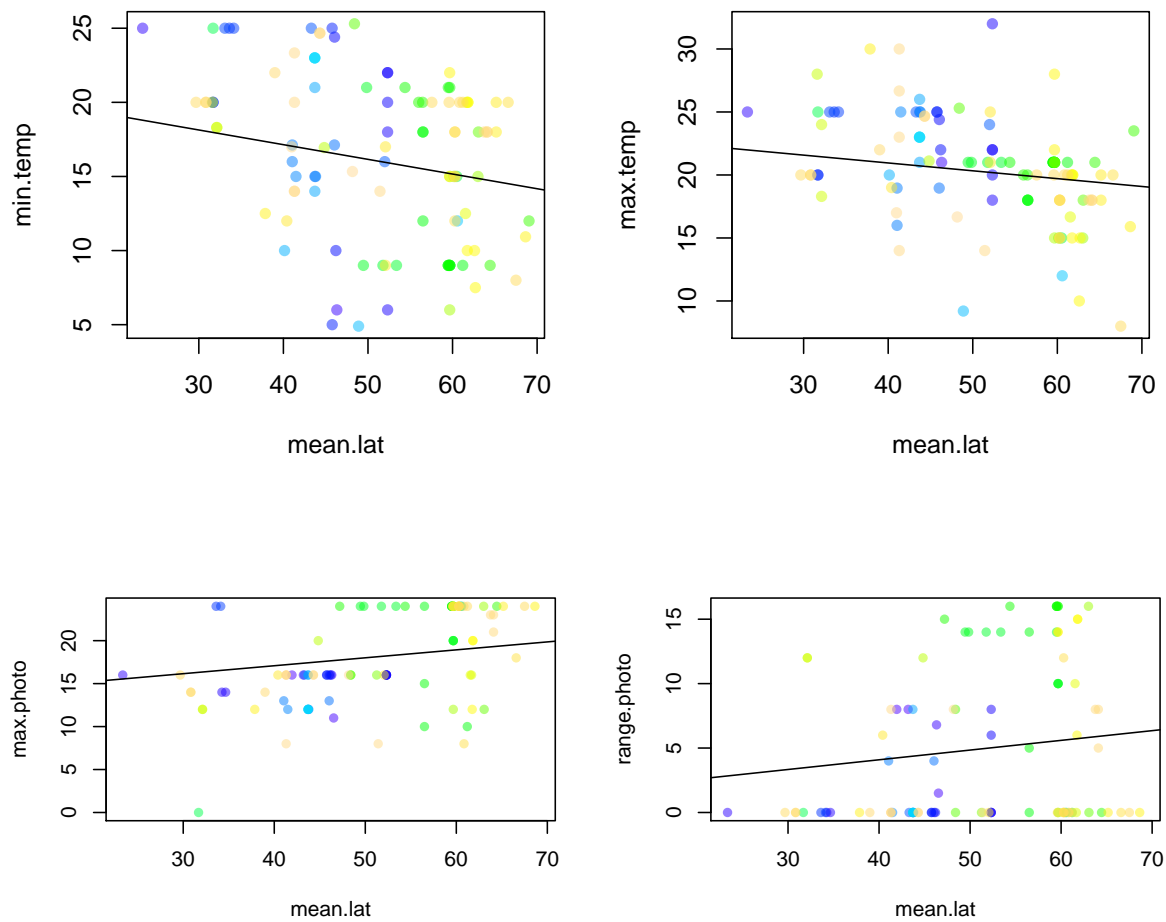Figure 3: Cues manipulated over time.

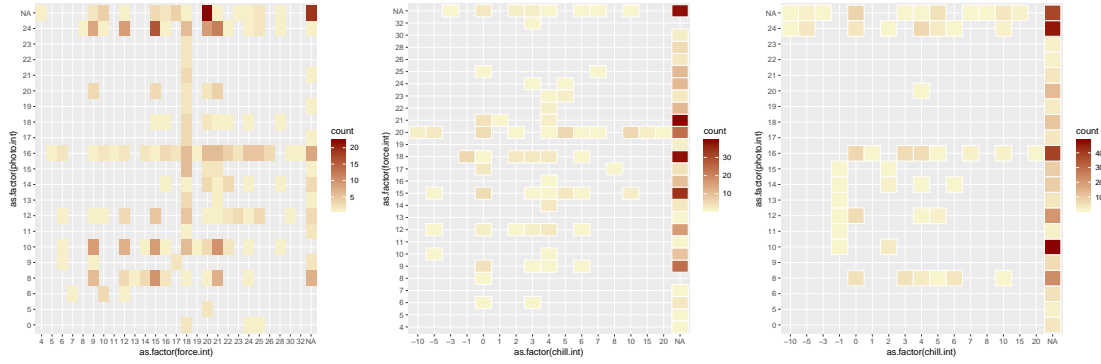Figure 4: One correlation with latitude plot? Or more?

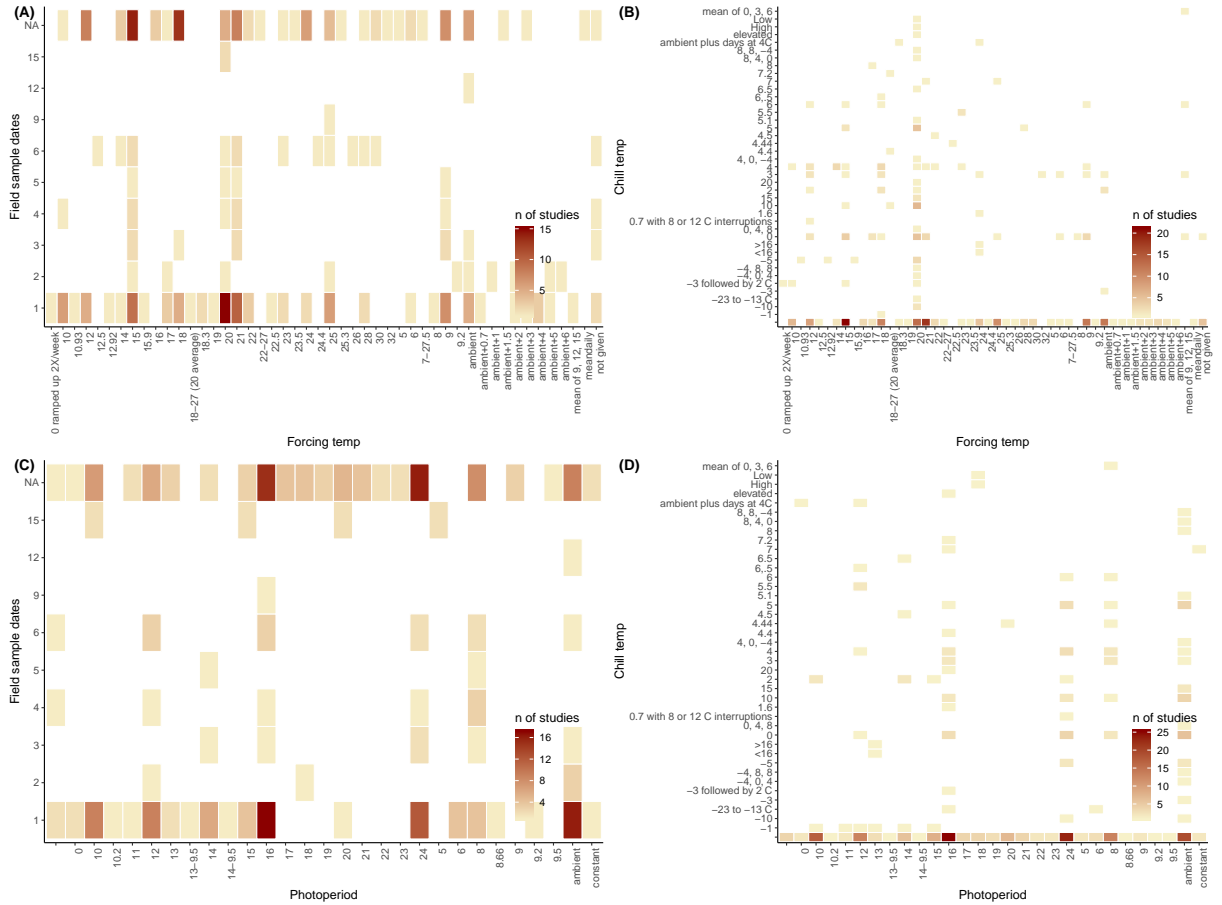Figure 5: Our old heat maps, based only on what we could convert to numeric easily.

Figure 6: Heatmap using all woody species data in OSPREE: shows number of studies for different study designs (we could clean this up in various ways ... perhaps only show budbreak or budburst studies?).
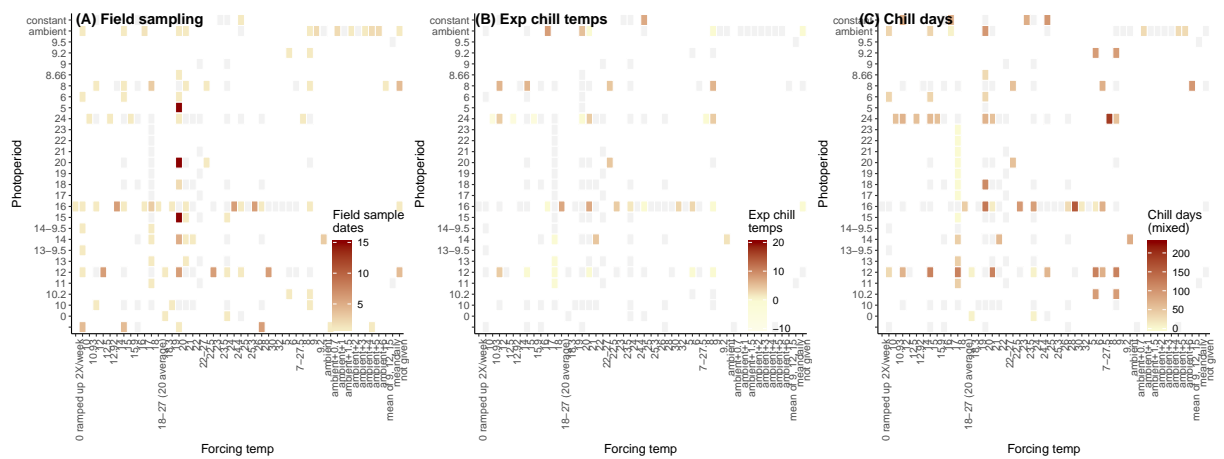
Figure 7: Heatmap using all woody species data in OSPREE and attempting to show three-way treatments (note that chilldays includes a mix of experimental chilling and field-sample day chilling)