

# Supplemental materials: Chilling dominates spring phenological responses to warming

A.K. Ettinger, C. Chamberlain, I. Morales-Castilla, D. Buonaiuto, D. F. B. Flynn, T. Savas, J. Samaha & E. M. Wolkovich

July 8, 2019

## Supplemental Methods

### Observed Spring Phenology Responses in Experimental Environments (OSPREE) database

We searched the literature for research papers which experimentally addressed controls of temperature, chilling, and/or photoperiod requirements on the spring phenology of woody plant species. To identify phenological experiments that manipulated chilling, forcing, and/or photoperiod, we searched both ISI Web of Science and Google Scholar in July 2015 with the following terms:

1. TOPIC = (budburst OR leaf-out) AND (photoperiod or daylength) AND temperature\*, which yielded 85 publications
2. TOPIC = (budburst OR leaf-out) AND dorman\*, which yielded 193 publications

The initial searches yielded 201 papers, which we reviewed and assessed for inclusion in the database using the following criteria: focusing on woody plants in temperate ecosystems, and testing for at least photoperiod or temperature effects on budburst, leafout or flowering.

Although most all studies measured days to budburst, many communicated results differently (e.g. days to budburst, degree-days to budburst, percent budburst, number of leaves etc). We standardized papers to common units whenever possible (details below) and further restricted studies to those for which forcing, chilling, and photoperiod treatments could be quantitatively identified. For this paper, we focus on studies measuring days to budburst. This subset of OSPREE includes data across 66 studies, 39 years, and 203 species (Fig. S1).

Some species are only represented in one dataset in the OSPREE database, making it impossible to statistically differentiate between species, study, and treatment effects for these taxa. To address this, we combined species found in only one study into “complexes” at the level of genera—such that each taxonomic unit we use in our model occurs across multiple studies (and treatments). Thus our taxonomic units of analysis are “species complexes,” which are either species represented in >1 dataset or complexes combining multiple species within a genus that are each singly represented in the dataset. Species represented in only one dataset with no congeners in other datasets were excluded from our analysis.

*Defining budburst*

Most studies defined budburst as initial “green tips” (33/49 studies). Select studies defined budburst as a specific increment of growth (e.g., “0.5cm of new growth”) or as bud swell, leaf emergence, leaf unfolded, open bud scales, or petiole emerged. The remaining studies (4/49) did not include a definition of budburst. The majority of studies using the above definitions (34/49) required only one bud to have met the defined criteria of budburst, however, the remaining studies implemented specific thresholds to be met (i.e., 10-100% of all buds on an individual needed to have bursted bud). We only included studies with at least 49.5% budburst. For studies with multiple measurements of percent budburst over time, we used the days to budburst when percent budburst was closest to 90%.

#### *Estimating chilling*

Chilling was reported far less in the OSPREE database than forcing and photoperiod. While not all studies applied multiple treatments of forcing and/or photoperiod they generally all maintained and explicitly defined their forcing temperatures and daylengths. In contrast, we found that most studies did not experimentally apply chilling by manipulating duration or temperature of chilling in controlled environments, nor did most quantify the total chilling imposed in their experiment. We therefore calculated the total chilling imposed by all studies, as it would otherwise have been impossible to provide estimates with only experimental chilling given the rarity of such study designs (Fig S2).

To estimate total chilling we combined chilling from the field (i.e., chilling before plant material was brought into environmental growth chambers) and experimental chilling (i.e., chilling plant material experienced in environmental growth chambers) into two widely used metrics of chilling: Utah units and dynamic chill portions (Dennis, 2003). We used the *chillR* package (version 0.70.17) in R (R Development Core Team, 2017; Luedeling, 2019), version 3.6.0 to calculate both positive Utah units and dynamic chill portions. Utah chilling units accumulated the most at temperatures between 2.4-9.1°C but slightly less at temperatures between 1.4-2.4°C and from 9.1-12.4°C. Utah units were reduced when temperatures fell below or exceeded this range. Chill portions accumulated when temperatures were between 0 and 7.2°C. We note that these models for chilling (both of which were developed for peach species) are *hypotheses* for how chilling may accumulate to affect the process of endodormancy release, but are likely to be inaccurate for many species. These models are, however, some of our current best approximations. We found the effects of chilling and other cues remain qualitatively consistent across the two metrics of totalchilling, though total chilling and photoperiod estimates were slightly lower using chill portions compared to Utah units (Table S2).

#### *Estimating forcing & photoperiod*

Our studies included a diversity of designs for applying forcing and/or photoperiod experimentally, including studies that imposed constant forcing temperatures and forcing temperatures that varied between day and night. Additionally several studies applied forcing or photoperiod using a “ramped” design, such that treatments increased or decreased gradually over time throughout the duration of the application. For all studies we used the daylength of light as our photoperiod estimate (e.g., a study with 8 hours of light and 16 hours of dark was recorded as ‘8’). For forcing, we used the temperature applied, when forcing temperatures were (i.e., the same temperature was applied 24 hours per day). If forcing varied with photoperiod, we estimated the mean daily temperature weighted by the hours that temperature was applied. Similarly, for ramped studies, we calculated a weighted average of forcing temperature over the period from when forcing treatments were applied until budburst day

## Models

We fit three overall models: the main budburst model, fit to all studies in OSPREE that measured days to budburst; the latitude model, which included only studies that had provenance latitude information, and a model to examine how the design of chilling treatments affects estimated effects. Given the complexity of our meta-analytic data we fit each model separately, and present the main model in the main text as it was designed to best estimate chilling, forcing and photoperiod cues (our primary goal here). The other

two models represent subsets of the data in the main model that allow more direct tests of relevant, related questions.

As our primary goal was to directly compare the effects of chilling, forcing and photoperiod we standardized these predictor variables (Gelman and Hill, 2006). We followed well-established methods methods of subtracting the mean and dividing by the standard deviation (Gelman and Hill, 2006) to yield “z-score” values for all predictor variables (total chilling units, forcing temperatures, and photoperiods in the experiments).

The models were fit using the programming languages **Stan** (Carpenter et al., 2017)([www.mc-stan.org](http://www.mc-stan.org)), accessed via the *rstan* package (version 2.18.0) in R (R Development Core Team, 2017; Stan Development Team, 2018), version 3.6.0. Stan provides efficient MCMC sampling via a No-U-Turn Hamiltonian Monte Carlo approach (more details can be found in Gelman et al. (2014) and in Carpenter et al. (2017)). We validated that our models using test data, then fit the following models:

1. Main budburst model:

$$y_i = N(\alpha_{sp[i]} + \beta_{forcing_{sp[i]}} + \beta_{photoperiod_{sp[i]}} + \beta_{chilling_{sp[i]}} + \epsilon_i, \epsilon_i \sim N(0, \sigma_y^2)$$

The  $\alpha$  and each of the three  $\beta$  coefficients were modeled at the species level, as follows:

$$\begin{aligned} \alpha_{sp} &\sim N(\mu_\alpha, \sigma_\alpha) \\ \beta_{forcing_{sp}} &\sim N(\mu_{forcing}, \sigma_{forcing}) \\ \beta_{photoperiod_{sp}} &\sim N(\mu_{photoperiod}, \sigma_{photoperiod}) \\ \beta_{chillingsp} &\sim N(\mu_{chilling}, \sigma_{chilling}) \end{aligned}$$

2. Latitude model: Given continuing debate over the role of photoperiod on budburst timing across a species’ latitudinal range (e.g., Zohner et al., 2016; Gauzere et al., 2017), we examined the effect of including latitude in a model similar to our main one, but designed to estimate latitude effects. This model estimated the effects of each phenological cue (chilling, forcing, photoperiod) on days to budburst (as in the main model), in addition to the effect of latitude and the interaction of photoperiod by latitude. We include this interaction because photoperiod effects are expected to vary by latitude and this interaction may have important implications under climate change (Saikkonen et al., 2012; Way and Montgomery, 2015; Gauzere et al., 2017).

We followed the guidelines above for including species or species complex (see *Observed Spring Phenology Responses in Experimental Environments (OSPREE) database* section above), then subsetted the species and species complexes to include only those that had multiple provenance locations. This yielded the following model:

$$\begin{aligned} y_i = N(\alpha_{sp[i]} + \beta_{forcing_{sp[i]}} + \beta_{photoperiod_{sp[i]}} + \beta_{chilling_{sp[i]}} + \beta_{latitude_{sp[i]}} \\ + \beta_{photoperiod:latitude_{sp[i]}} + \epsilon_i, \epsilon_i \sim N(0, \sigma_y^2) \end{aligned}$$

The  $\alpha$  and each of the five  $\beta$  coefficients were modeled at the species level, as follows:

$$\begin{aligned} \alpha_{sp} &\sim N(\mu_\alpha, \sigma_\alpha) \\ \beta_{forcing_{sp}} &\sim N(\mu_{forcing}, \sigma_{forcing}) \\ \beta_{photoperiod_{sp}} &\sim N(\mu_{photoperiod}, \sigma_{photoperiod}) \\ \beta_{chillingsp} &\sim N(\mu_{chilling}, \sigma_{chilling}) \\ \beta_{latitudesp} &\sim N(\mu_{latitude}, \sigma_{latitude}) \\ \beta_{photoperiod:latitude_{sp}} &\sim N(\mu_{photoperiod:latitude}, \sigma_{photoperiod:latitude}) \end{aligned}$$

3. Chilling study design model: As we found chilling to be the strongest cue, and given how few studies directly manipulate it (Fig S2), we also used a subset of our data to estimate how a study's experimental design for chilling impacts model estimates. For this, we included only species or species complexes used in both experiments that employed the Weinberger method (in this method plant tissue is sequentially removed from the field followed and then exposed to 'forcing' conditions, with the assumption that tissues collected later experience more field chilling Weinberger et al., 1950) and those that experimentally manipulated chilling (i.e., by varying chilling temperatures and or the duration of chilling conditions). We defined Weinberger studies as those with two or more field sample dates, each two or more weeks apart, that did not otherwise manipulate chilling. The chilling study-design model was thus:

$$y_i = N(\alpha_{sp[i]} + \beta_{forcing} + \beta_{photoperiod} + \beta_{chilling} + \beta_{chillmethod} + \beta_{forcing:chillmethod}\beta_{photoperiod:chillmethod} + \beta_{chilling:chillmethod}\epsilon_i, \epsilon_i \sim N(0, \sigma_y^2)$$

$$\alpha_{sp} \sim N(\mu_\alpha, \sigma_\alpha)$$

For all models, we chose weakly informative priors; increasing the priors three-fold did not change the model results.

We ran four chains simultaneously, each with 1 500 warm-up iterations followed by 2 500 sampling iterations, yielding 4 000 posterior samples for each parameter. We assessed model performance through  $\bar{R}$  close to 1 and high  $n_{eff}$  (4 000 for most parameters, but as low as 713 for a few parameters in the latitude model), as well as visual consideration of chain convergence and posteriors (Gelman et al., 2014).

In our figures we show means  $\pm$  50% credible intervals from our models (Fig. 1, ??, S4), because our focus here is on the most likely value for each parameter (e.g., estimated response to forcing) and because they are computationally stable (Gelman et al., 2014; Carpenter et al., 2017). See Tables S2- S5 for 95% credible intervals.

#### *Meta-analytic modeling limitations*

As our focus is on experiments, which—by design—often impose high variation in phenological cues, we expected a linear model for chilling, forcing, and photoperiod would be most appropriate. Non-linear models, however, are often most appropriate for phenological cues, especially in nature where chilling may always be very high or extremely short photoperiods are rarely experienced. Thus we tested a non-linear (sigmoidal) model on the OSPREE data (citePMP users manual). As chilling was the least experimentally manipulated in our database, we examined whether a sigmoidal curve for chilling would be more appropriate, but found that it was a poorer fit than a comparable all-linear model ( $R^2$  = 0.53 versus 0.57), did not dramatically alter estimates of forcing (-0.83 versus -0.85) or photoperiod (-0.25 versus -0.13) and led to non-biologically relevant estimates of chilling. Fitting non-linear models to experimental data may require more data, and/or data at very high and low chilling, forcing and photoperiod values, than currently available.

An ideal model to predict budburst would potentially include (but is not limited to): interactions between cues, sigmoidal or other non-linearities to assess potential threshold effects, provenance location, methodological details (e.g., if tissue was seedlings versus twigs, or whether temperatures were constant or varied each day, etc). As with all models, though, we were limited in how many parameters we could estimate given available data. Thus we focused on species differences and used additional models to assess some of the potentially largest other effects (latitude, methods of estimating chilling). We were unable to estimate interactions between cues in our meta-analysis because very few studies design experiments to test for interactions between chilling, forcing, and photoperiod (cite table with number of interactions from coding challenge!). The few that do incorporate interactions generally use the Weinberger method, which is not designed to robustly tease out of the effects of multiple cues (Table S5, Fig. S4). Similarly we found variation in study

material/tissue and variation in thermoperiodicity was too infrequent to test for effects with current data. Our estimated therefore effects average over interactions (Gelman and Hill, 2006), but identifying them in future research will be critical to understanding and predicting budburst.

## Applying our model to Central European data

.. Our results integrate over a large range of chilling, forcing and photoperiod conditions (e.g., forcing treatment temperatures—as defined by the each study’s authors—in growth chamber experiments ranged from 0-32 °C and chilling temperatures ranged from -10-16 °C). We also wished to understand how our findings may apply to conditions more commonly found in nature, where conditions may vary dramatically from those applied in chamber experiments. For example, low amounts of total chilling can be applied in experiments compared to natural chilling found in many temperate areas. Additionally, chilling temperature and total chilling are more correlated in nature than in experimental conditions (?).

Our goal was to understand the implications of our model estimates on budburst across varying degrees of warming, given natural daily and seasonal cycles in the climate of a well-studied area. We therefore focused on defined levels of warming given historical climate. We use this approach instead of climate change projections from Global Circulation Models for several reasons: First and foremost, this approach allowed us to focus ... [LIZZIE WILL FINISH THIS LATER] which we would have first had to understand how those projections related to degrees of warming). We emphasize that our applications are unlikely to be accurate forecasts of budburst dates, even for the locations for which we use historical climate and budburst data. They provide instead an example of ways in which natural conditions can differ from experimental conditions, and the implications of these differences for budburst.

To apply our model to central Europe, we selected sites in Europe where temperature and budburst have been monitored since the 1950s (the Pan European Phenology Project, <http://www.pep725.eu>, PEP), including for two common European species that are prevalent in the OSPREE database: *Betula pendula* (silver birch) and *Fagus sylvatica*. We used a European-wide gridded climate dataset (Cornes et al., 2018) to extract daily minimum and maximum temperature for the grid cells where observations for these two species were available. We extracted temperature data from 1951 through 1961 (selected as a pre-warming time period) and used these data to estimate annual values for total winter chilling (from 1 September through 30 April, in Utah units, using chillR) and mean spring forcing (mean temperature from 1 March through 30 April). We used our main model to estimate budburst under these pre-warming conditions, and then investigated model predictions of budburst given different levels of warming (from 1-7 °C) above this baseline, including altered total chilling and forcing estimates, as well as potential declines in photoperiod due to advancing phenology. To estimate shifts in photoperiod with warming, we first estimated the shift in days to budburst with specified warming (i.e., due to chilling and forcing). We used this estimate to identify the change in photoperiod between the day of year during the pre- and post-warming periods and then re-fit the model with this new photoperiod.

We applied our model at all latitudes and longitudes included in the PEP database between 1951 and 1961 for *Betula pendula*. We selected two of these sites for *Betula pendula*, as well as two sites where *Fagus sylvatica* occurs, in order to compare budburst responses across species that differ in their responses to chilling, forcing, and photoperiod, as well as sites that differ in baseline climate (Figures 3, S6).

Cat- please add a paragraph here describing more about the comparison of your

## Supplemental Discussion

### Surprising species-specific responses

For a few taxa, estimates for forcing were positive; these were *Fagus grandifolia*, *Acer*-complex, *Fraxinus* complex, *Cornus alba*. We interpret these as...

### Potential statistical artifacts in declines of temperature sensitivity in observational long-term data

As our model results do not predict a dramatic decline in temperature sensitivity in Central Europe, as has been observed (e.g., Fu et al., 2015), we tested whether observed declines could instead be due to a statistical artifact. Researchers today commonly estimate temperature sensitivity via a linear regression of annual budburst date versus mean or other aggregated metrics of spring temperature yielding estimates in days/ $^{\circ}\text{C}$ . However, if warming produces systematically warmer daily temperatures this method will inherently estimate lower sensitivities with warming, because the ‘days’ unit will effectively have increased in the thermal sum it represents (that is the unit of ‘days’ is non-stationary in recent decades).

To test this hypothesis we compared observed trends with simple simulations. First, we collated PEP 725 data (Templ et al., 2018) for *Betula pendula* for all sites with leafout data each year from two 10-year time-periods: a period before significant anthropogenic warming (1951-1960) and a period with significant warming (2001-2010, see IPCC, 2014). We used leafout data (BBCH=11; which is defined as “leaf unfolding (first visible leaf stalk)” in the PEP725 database) instead of budburst (BBCH=7; defined as “Beginning of sprouting”) as leafout data are far more common in the PEP 725 database. Next, we simulated budburst data with constant cues. For this, we did not include any chilling or photoperiod cues, but assumed budburst occurred after a certain thermal sum, estimated via growing degree days with a base temperature of 0 $^{\circ}\text{C}$ . We then estimated temperature sensitivity (days/ $^{\circ}\text{C}$ ) and the difference in these estimates given different levels of spring warming. For the simulations shown here we used a GDD (growing degree day) requirement of 150, a base mean spring temperature of 6 $^{\circ}\text{C}$  with a variance of 3 $^{\circ}\text{C}$ , and estimated temperature sensitivity for 10-year periods for 45 simulated sites (these values were chosen to best match the PEP 725 data, but note that the general findings are robust to other combinations of these parameter values).

As expected temperature sensitivity estimates for *Betula pendula* from PEP 725 declined across the two time periods in step with warming. Across the sites studied here we estimated a decline of 0.8 $\pm$ 0.3 days/ $^{\circ}\text{C}$  (comparing 2001-2010 and 1951-1960) and 1.1 $\pm$ 0.2 $^{\circ}\text{C}$  warming; this estimate was very similar to simulations given constant cues and 1 $^{\circ}\text{C}$  warming (Fig. S7).

Additionally, several other metrics suggest declines may be more statistical than biological. Research suggests substantial declines in chilling that could lead to observed shifts in sensitivity to warm should increase variance in leafout timing (Ford Kevin et al., 2016). In contrast, in both the real and simulated data variance in leafout date declined over time—this would be expected if plants use a thermal sum threshold of forcing to leaf out and warming produces systematically warmer days. In the PEP 725 data we found a decline in leafout variance of 58% (in recent years, compared to earlier years), compared to a decline of 37% in the simulations. Additionally we found little change in accumulated chilling (1 September - 1 March of each year) in the PEP 725 data across the two time points (2247 $\pm$ 31 Utah units in 1951-1960, compared to 2236 $\pm$ 20 Utah units in 2001-2010), further suggesting that shifts in chilling do not explain the declining sensitivities. Simple plots of the chilling and forcing required for budburst suggest very low chilling is often required to dramatically increase the forcing required for budburst (Fig. fig:pepgddchill).

This potential artifact adds to existing research that has documented the statistical challenges of accurately

estimating temperature sensitivities from long-term data (Güsewell et al., 2017; Clark et al., 2014) and may be overcome by some methods. Research that measures sensitivity as a thermal sum or other temperature metric (e.g., GDD) until leafout should be less vulnerable to the artifact. Indeed, the PEP 725 data we found little difference across the two time-periods in GDD ( $68.7 \pm 2.6$  in 1950-1960 versus  $61.5 \pm 2.0$  in 2000-2010 for GDD calculated from January 1st to leafout with a base temperature of  $0^{\circ}\text{C}$ ; and a mean temperature in the 30 days before leafout of  $6.8^{\circ}\text{C} \pm 0.1$  in 1950-1960 versus  $6.6^{\circ}\text{C} \pm 0.1$ ). This method, however, is also vulnerable to other issues: as researchers must select the day to start accumulating or averaging temperatures it should work best when this day is always after endodormancy break—when plants are most responsive to forcing (Chuine et al., 2016). As climate change may push endodormancy break later and later in some regions, this method could inaccurately attribute changes in other cues to shifts in forcing (Güsewell et al., 2017). Without measures of endodormancy break (Chuine et al., 2016), we suggest efforts to accurately estimate cues from long-term observational data may be difficult to impossible without additional physiological information from controlled environment experiments.

## References

- Carpenter, B., A. Gelman, M. Hoffman, D. Lee, B. Goodrich, M. Betancourt, M. A. Brubaker, J. Guo, P. Li, and R. Allen. 2017. Stan: A probabilistic programming language. *Journal of Statistical Software* 76:10.18637/jss.v076.i01.
- Chuine, I., M. Bonhomme, J.-M. Legave, I. García de Cortázar-Atauri, G. Charrier, A. Lacointe, and T. Améglio. 2016. Can phenological models predict tree phenology accurately in the future? the unrevealed hurdle of endodormancy break. *Global change biology* 22:3444–3460.
- Clark, J. S., J. Melillo, J. Mohan, and C. Salk. 2014. The seasonal timing of warming that controls onset of the growing season. *Global Change Biology* 20:1136–1145.
- Cornes, R. C., G. van der Schrier, E. J. van den Besselaar, and P. D. Jones. 2018. An ensemble version of the e-obs temperature and precipitation data sets. *Journal of Geophysical Research: Atmospheres* 123:9391–9409.
- Dennis, F. 2003. Problems in standardizing methods for evaluating the chilling requirements for the breaking of dormancy in buds of woody plants. *HortScience* 38:347–350.
- Ford Kevin, R., A. Harrington Constance, S. Bansal, J. Gould Peter, and J. B. St. Clair. 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., H. F. Zhao, S. L. Piao, M. Peaucelle, S. S. Peng, G. Y. Zhou, P. Ciais, M. T. Huang, A. Menzel, J. P. Uelas, Y. Song, Y. Vitasse, Z. Z. Zeng, and I. A. Janssens. 2015. Declining global warming effects on the phenology of spring leaf unfolding. *Nature* 526:104–107.
- Gauzere, J., S. Delzon, H. Davi, M. Bonhomme, I. G. de Cortazar-Atauri, and I. Chuine. 2017. Integrating interactive effects of chilling and photoperiod in phenological process-based models. A case study with two European tree species: *Fagus sylvatica* and *Quercus petraea*. *Agricultural and Forest Meteorology* 244:9–20.
- Gelman, A., J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin. 2014. *Bayesian Data Analysis*. 3rd ed. CRC Press, New York.
- Gelman, A., and J. Hill. 2006. *Data analysis using regression and multilevel/hierarchical models*. Cambridge university press.

- Güsewell, S., R. Furrer, R. Gehrig, and B. Pietragalla. 2017. Changes in temperature sensitivity of spring phenology with recent climate warming in switzerland are related to shifts of the preseason. *Global change biology* 23:5189–5202.
- IPCC. 2014. Climate Change 2014: Impacts, Adaptation, and Vulnerability. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.
- Luedeling, E. 2019. Statistical methods for phenology analysis in temperate fruit .
- R Development Core Team. 2017. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
- Saikkonen, K., K. Taulavuori, T. Hyvönen, P. E. Gundel, C. E. Hamilton, I. Vänninen, A. Nissinen, and M. Helander. 2012. Climate change-driven species' range shifts filtered by photoperiodism. *Nature Climate Change* 2:239.
- Stan Development Team. 2018. RStan: the R interface to Stan. R package version 2.17.3.
- Templ, B., E. Koch, K. Bolmgren, M. Ungersböck, A. Paul, H. Scheifinger, T. Rutishauser, M. Busto, F.-M. Chmielewski, L. Hájková, S. Hodžić, F. Kaspar, B. Pietragalla, R. Romero-Fresneda, A. Tolvanen, V. Vučetić, K. Zimmermann, and A. Zust. 2018. Pan european phenological database (pep725): a single point of access for european data. *International Journal of Biometeorology* 62:1109–1113.
- Way, D. A., and R. A. Montgomery. 2015. Photoperiod constraints on tree phenology, performance and migration in a warming world. *Plant, Cell & Environment* 38:1725–1736.
- Weinberger, J. H., et al. 1950. Chilling requirements of peach varieties. Pages 122–28 in Proceedings. American Society for Horticultural Science. Vol. 56.
- Zohner, C. M., B. M. Benito, J. C. Svenning, and S. S. Renner. 2016. Day length unlikely to constrain climate-driven shifts in leaf-out times of northern woody plants. *Nature Climate Change* .

## Supplemental Tables

Still need:

1. A table that goes with Figure 3 (the 4-paneled 3D forecasting figure) that includes the mean budburst day of year, chilling estimates, mean winter temperature, and spring (forcing) temperatures for the 4 sites with no warming and with the 7 levels of warming.
2. The count interactions table?

## Supplemental Figures

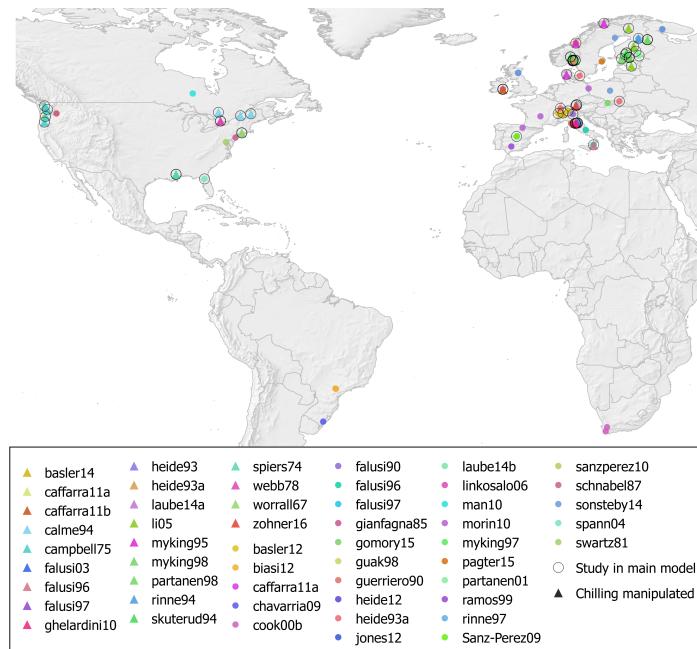


Figure S1: Map of days to budburst experiments in the OSPREE database.

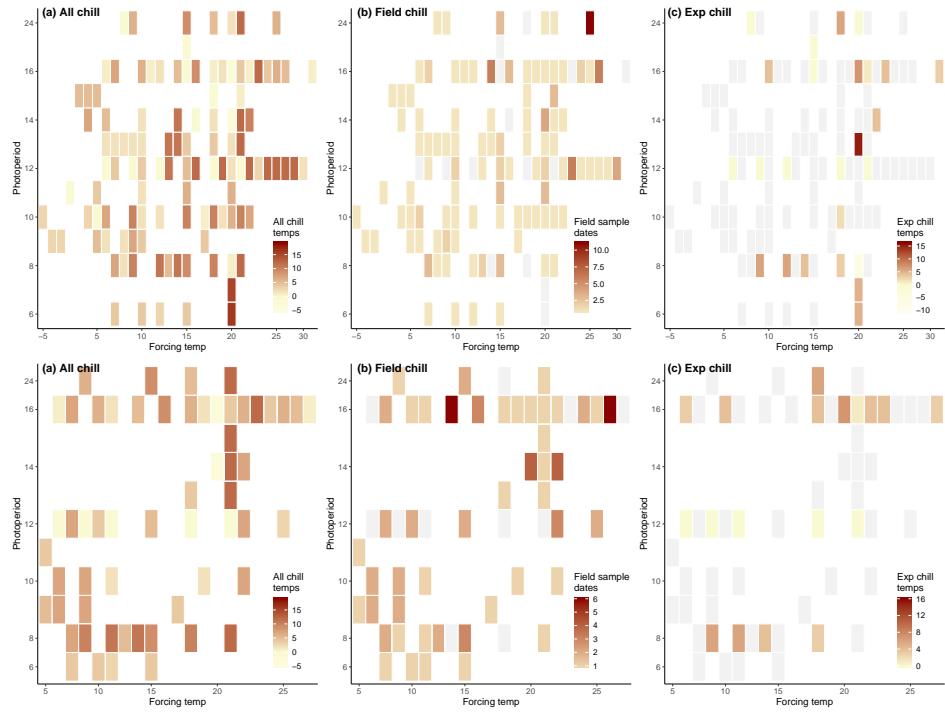


Figure S2: **Heatmaps of treatments** top row shows full dataset, bottom row shows main model

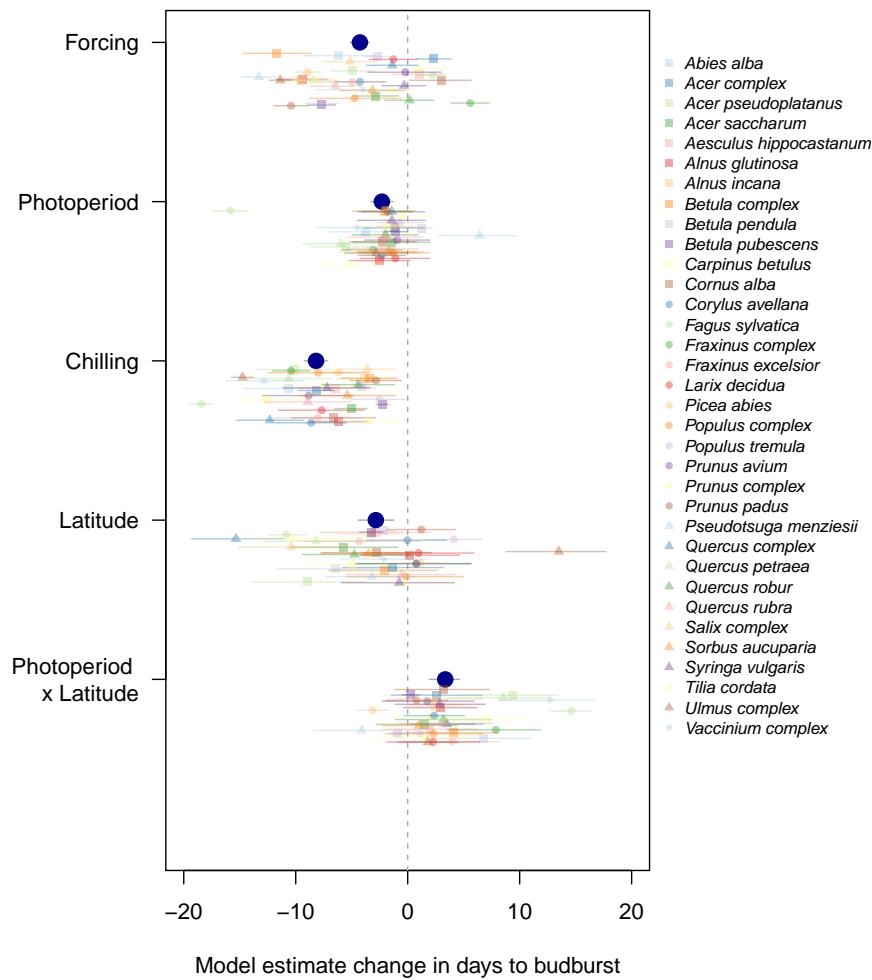


Figure S3: Estimates for effects of chilling exceeded estimates for forcing, photoperiod, latitude, and the interaction between latitude and photoperiod, for most species, in the latitude budburst model fit to centered data, including the subset of studies in OSPREE database that XXX. Here we show estimates from the model fit to centered data, enabling comparisons of effects sizes across predictors, and using Utah units to quantify chilling.

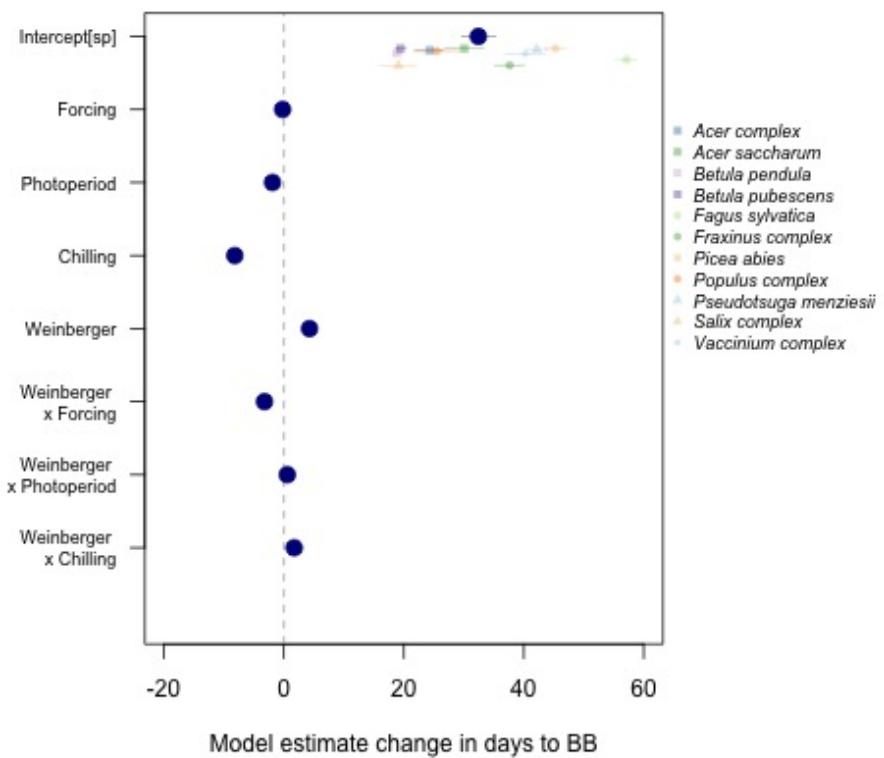


Figure S4: Comparison of estimated effects for environmental parameters for overlapping species included in both Weinberger method studies and non-Weinberger method studies. The effect of chilling is estimated to be weaker, and the effect of forcing stronger in Weinberger studies.

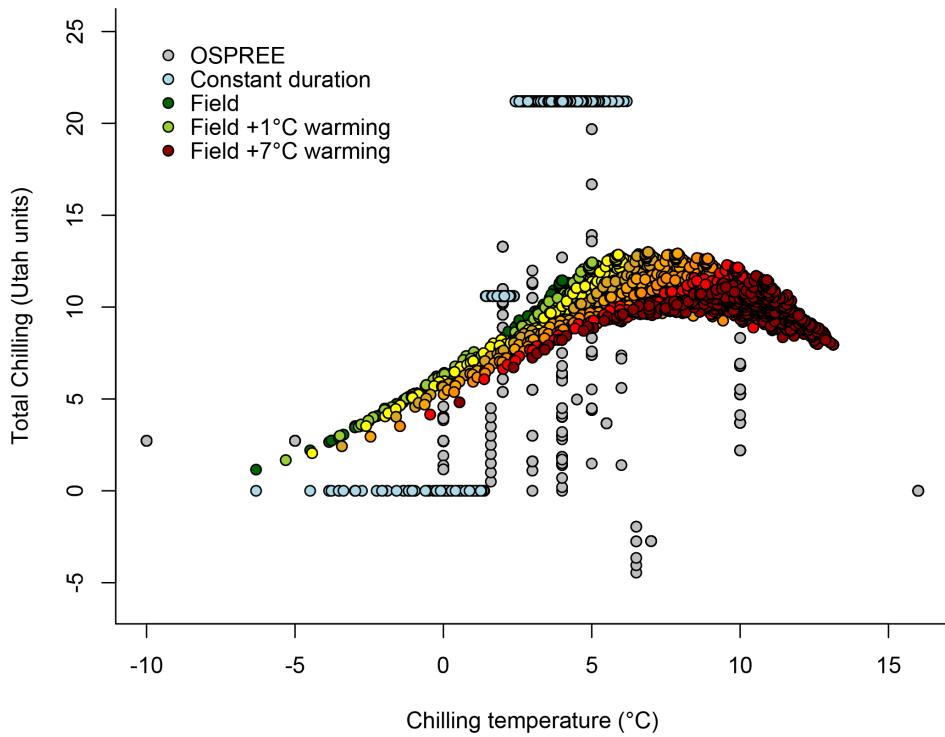


Figure S5: Chilling accumulates differently in experiments with constant temperatures versus natural systems in which temperature is more strongly correlated with chilling.

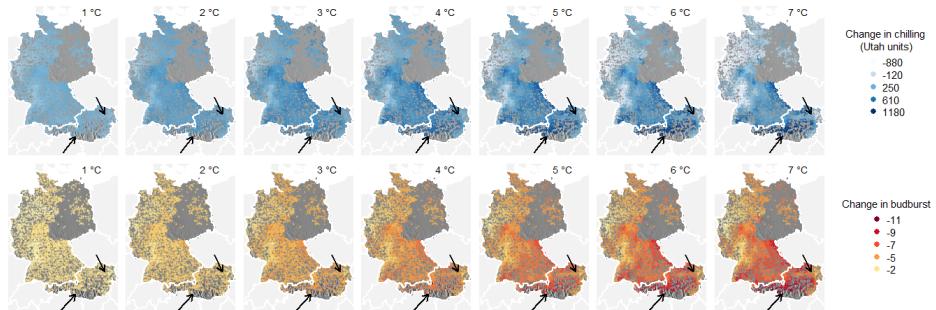


Figure S6: Forecasted changes in chilling (top panel) and leafout for *Betula pendula* (bottom panel), in locations included in the PEP database, where phenology dates are known for the pre-warming time period (1951-1960). Changes in chilling and budburst are calculated relative to the mean chilling and budburst dates during this pre-warming time period for each location. Arrows indicate sites shown in Figure 3A and B.

Table S1: Species included in the OSPREE database.

spname	stnum	studies
<i>Abies.alba</i>	2	basler12,laube14a
<i>Abies.homolepis</i>	1	laube14a
<i>Acer.barbinerve</i>	1	zohner16
<i>Acer.campestre</i>	1	zohner16
<i>Acer.ginnala</i>	1	zohner16
<i>Acer.negundo</i>	1	laube14a
<i>Acer.platanoides</i>	1	zohner16
<i>Acer.pseudoplatanus</i>	3	basler12,basler14,laube14a
<i>Acer.saccharinum</i>	1	webb78
<i>Acer.saccharum</i>	3	calme94,laube14a,webb78
<i>Acer.tataricum</i>	1	laube14a
<i>Actinidia.deliciosa</i>	2	biasi12,guerrero90
<i>Aesculus.flava</i>	1	zohner16
<i>Aesculus.hippocastanum</i>	3	basler12,laube14a,zohner16
<i>Aesculus.parviflora</i>	1	zohner16
<i>Alnus.glutinosa</i>	2	heide93,myking98
<i>Alnus.incana</i>	2	heide93,zohner16
<i>Alnus.maximowiczii</i>	1	zohner16
<i>Amelanchier.alnifolia</i>	1	zohner16
<i>Amelanchier.florida</i>	1	zohner16
<i>Amelanchier.laevigata</i>	1	zohner16
<i>Amorpha.fruticosa</i>	1	laube14a
<i>Aronia.melanocarpa</i>	1	zohner16
<i>Berberis.dielsiana</i>	1	zohner16
<i>Betula.alleganiensis</i>	1	calme94
<i>Betula.lenta</i>	1	zohner16
<i>Betula.nana</i>	1	zohner16
<i>Betula.pendula</i>	10	heide93,li05,rinne97,basler12,laube14a,laube14b,linkosal06,myking95,myking95,zohner16
<i>Betula.populifolia</i>	1	zohner16
<i>Betula.pubescens</i>	6	heide93,rinne94,caffarra11a,caffarra11b,myking95,myking97
<i>Buddleja.albiflora</i>	1	zohner16
<i>Buddleja.alternifolia</i>	1	zohner16
<i>Buddleja.davidii</i>	1	zohner16
<i>Caragana.pygmaea</i>	1	zohner16
<i>Carpinus.betulus</i>	3	heide93a,laube14a,zohner16
<i>Carpinus.laxiflora</i>	1	zohner16
<i>Carpinus.monbeigiana</i>	1	zohner16
<i>Carya.cordiformis</i>	1	zohner16
<i>Carya.laciniosa</i>	1	zohner16
<i>Carya.ovata</i>	1	zohner16
<i>Castanea.sativa</i>	1	zohner16
<i>Cedrus.libani</i>	1	zohner16
<i>Celtis.caucasica</i>	1	zohner16
<i>Celtis.laevigata</i>	1	zohner16
<i>Celtis.occidentalis</i>	1	zohner16
<i>Cephalanthus.occidentalis</i>	1	zohner16
<i>Cercidiphyllum.japonicum</i>	1	zohner16
<i>Cercidiphyllum.magnificum</i>	1	zohner16
<i>Cercis.canadensis</i>	1	zohner16
<i>Cercis.chinensis</i>	1	zohner16
<i>Cladrastis.lutea</i>	1	zohner16
<i>Cornus.alba</i>	2	laube14a,zohner16
<i>Cornus.kousa</i>	1	zohner16
<i>Cornus.mas</i>	2	laube14a,laube14b
<i>Corylopsis.sinensis</i>	1	zohner16
<i>Corylopsis.spicata</i>	1	zohner16
<i>Corylus.avellana</i>	4	basler12,heide93,laube14a,zohner16
<i>Corylus.heterophylla</i>	1	zohner16
<i>Corylus.sieboldiana</i>	1	zohner16
<i>Decaisnea.fargesii</i>	1	zohner16
<i>Deutzia.gracilis</i>	1	zohner16
<i>Deutzia.scabra</i>	1	zohner16
<i>Elaeagnus.ebbingei</i>	1	zohner16
<i>Eleutherococcus.senticosus</i>	1	zohner16
<i>Eleutherococcus.setchuenensis</i>	1	zohner16
<i>Euonymus.europaeus</i>	1	zohner16
<i>Euonymus.latifolius</i>	1	zohner16
<i>Fagus.crenata</i>	1	zohner16
<i>Fagus.engleriana</i>	1	zohner16
<i>Fagus.orientalis</i>	1	zohner16
<i>Fagus.sylvatica</i>	10	falus190,falus196,falus197,basler12,basler14,caffarra11a,heide93a,heide93a,zohner16

Table S2: **Estimates from model fit with standardized predictors.** The model we present in the main text uses Utah units for chilling and includes studies that experimentally manipulated forcing and photoperiod. Using instead a model with chilling in Chill Portions results in quantitatively different species-level and overall estimates, though the results are qualitatively similar. We also present coefficients from a model including all species (i.e., with crops) with all treatment types. We present posterior means, as well as 25th, 75th, 2.5th and 95th percentiles from models in which the predictors have been standardized so that they are directly comparable.

	Utah units					Chill portions					All species				
	mean	25%	75%	2.5%	97.5%	mean	25%	75%	2.5%	97.5%	mean	25%	75%	2.5%	97.5%
$\mu_\alpha$	29.94	28.77	31.1	26.45	33.29	30.73	29.52	31.97	27.07	34.41	30.89	30.14	31.61	28.71	33.19
$\mu_{forcing}$	-4.36	-5.12	-3.61	-6.6	-2.1	-4.83	-5.64	-4.06	-7.1	-2.47	-6.17	-7.02	-5.29	-8.86	-3.64
$\mu_{photoperiod}$	-3.15	-3.97	-2.3	-5.53	-0.74	-3.18	-3.92	-2.42	-5.4	-0.96	-1.02	-1.44	-0.61	-2.2	0.25
$\mu_{chilling}$	-8.89	-9.93	-7.81	-12.03	-5.8	-8.2	-9.27	-7.19	-11.18	-5.07	-8	-8.55	-7.45	-9.62	-6.4
$\sigma_\alpha$	9.41	8.51	10.18	7.19	12.31	10.18	9.2	11.05	7.76	13.08	14.37	13.71	15	12.63	16.3
$\sigma_{forcing}$	5.67	4.99	6.29	4.01	7.75	6.05	5.34	6.66	4.31	8.2	8.73	7.94	9.44	6.73	11.06
$\sigma_{photoperiod}$	5.24	4.4	5.95	3.32	7.87	4.47	3.83	5	2.93	6.54	3.68	3.35	3.97	2.79	4.71
$\sigma_{chilling}$	7.36	6.48	8.1	5.3	10.07	7.89	7.02	8.67	5.69	10.57	6.29	5.73	6.82	4.69	8.06
$\sigma_y$	15.77	15.59	15.96	15.24	16.31	15.47	15.27	15.65	14.94	16.01	14.94	14.8	15.07	14.56	15.33
$N_{sp}$	37					37					203				

Table S3: **Estimates from models fit with predictors on their natural scales,** so that effect sizes can be readily interpreted in a meaningful way (e.g., change in days of budburst per degree C of warming for forcing temperature). The model we present in the main text uses Utah units for chilling. Here we also present coefficients from a model included all species, including crops, and all treatment types. We present posterior means, as well as 25th, 75th, 2.5th and 95th percentiles, from models.

	Utah units					Chill portions					All species				
	mean	25%	75%	2.5%	97.5%	mean	25%	75%	2.5%	97.5%	mean	25%	75%	2.5%	97.5%
$\mu_\alpha$	62.87	60.21	65.53	54.87	70.84	66.94	63.87	69.99	57.95	75.87	62.7	61.05	64.36	57.82	67.74
$\mu_{forcing}$	-0.79	-0.91	-0.67	-1.16	-0.41	-0.85	-0.99	-0.72	-1.25	-0.46	-1.03	-1.12	-0.94	-1.29	-0.77
$\mu_{photoperiod}$	-0.54	-0.67	-0.41	-0.93	-0.17	-0.53	-0.66	-0.41	-0.91	-0.17	-0.14	-0.22	-0.07	-0.35	0.07
$\mu_{chilling}$	-2.84	-3.13	-2.53	-3.73	-1.97	-0.25	-0.28	-0.22	-0.33	-0.17	-2.48	-2.63	-2.34	-2.91	-2.08
$\sigma_\alpha$	19.16	17.35	20.79	14.53	24.78	21.97	19.93	23.74	16.86	28.42	17.7	16.81	18.54	15.33	20.38
$\sigma_{forcing}$	0.91	0.8	1.01	0.63	1.26	0.99	0.87	1.09	0.69	1.35	0.72	0.66	0.77	0.57	0.89
$\sigma_{photoperiod}$	0.79	0.67	0.88	0.51	1.16	0.7	0.6	0.79	0.46	1.03	0.59	0.54	0.64	0.45	0.75
$\sigma_{chilling}$	2.07	1.82	2.28	1.47	2.83	0.21	0.18	0.23	0.14	0.3	1.24	1.13	1.34	0.95	1.58
$\sigma_y$	15.82	15.63	16	15.27	16.37	15.52	15.34	15.7	15	16.08	15.16	15.02	15.3	14.78	15.57
$N_{sp}$	37					37					203				

Table S4: **Estimates from latitude model fit with standardized predictors.** Using a model with Utah chilling units and testing the effects of latitude plus the interaction between latitude and photoperiod results in slightly muted effects for forcing, photoperiod and chilling, though the results are qualitatively similar. We present posterior means, as well as 25th, 75th, 2.5th and 97.5th percentiles from models in which the predictors have been standardized so that they are directly comparable,

	mean	25%	75%	2.5%	97.5%
$\mu_\alpha$	29.13	27.85	30.49	25.09	32.93
$\mu_{forcing}$	-4.33	-5.15	-3.50	-6.80	-1.87
$\mu_{photoperiod}$	-2.28	-3.28	-1.31	-5.11	0.74
$\mu_{chilling}$	-8.18	-9.21	-7.15	-11.31	-5.05
$\mu_{latitude}$	-2.85	-4.34	-1.34	-7.43	1.60
$\mu_{photo:latitude}$	3.35	2.01	4.67	-0.62	7.42
$\sigma_\alpha$	8.86	7.77	9.80	6.23	12.34
$\sigma_{forcing}$	6	5.20	6.69	4.18	8.43
$\sigma_{photoperiod}$	5.34	4.45	6.09	3.33	8.11
$\sigma_{chilling}$	6.82	5.96	7.55	4.80	9.52
$\sigma_{latitude}$	8.1	6.44	9.50	4.21	13.30
$\sigma_{photo:latitude}$	6.79	5.37	7.95	3.63	11.19
$\sigma_y$	15.44	15.25	15.63	14.90	16.01
$N_{sp}$	36				

Table S5: **Estimates from Weinberger model fit with standardized predictors.** Using a model with Utah chilling units and testing the effects of the Weinberger method and the interaction between this method and the three main environmental cues show that budburst is generally later for Weinberger studies and the effect of chilling is muted while the effect of forcing is stronger. We present posterior means, as well as 25th, 75th, 2.5th and 97.5th percentiles from models in which the predictors have been standardized so that they are directly comparable.

	mean	25%	75%	2.5%	97.5%
$\mu_\alpha$	32.46	29.65	35.32	23.73	40.75
$\beta_{forcing}$	-0.21	-1.08	0.66	-2.75	2.39
$\beta_{photoperiod}$	-1.92	-2.50	-1.34	-3.65	-0.31
$\beta_{chilling}$	-8.22	-8.76	-7.68	-9.80	-6.61
$\sigma_\alpha$	13.34	11.28	14.96	8.81	20.24
$\sigma_y$	20.58	20.23	20.91	19.64	21.55
$\beta_{chillmethod}$	4.24	3.09	5.40	0.93	7.59
$\beta_{chilling:chillmethod}$	1.74	0.35	3.15	-2.43	5.73
$\beta_{forcing:chillmethod}$	-3.24	-4.35	-2.15	-6.50	-0.03
$\beta_{photoperiod:chillmethod}$	0.63	-0.42	1.67	-2.39	3.68
$N_{sp}$	11				

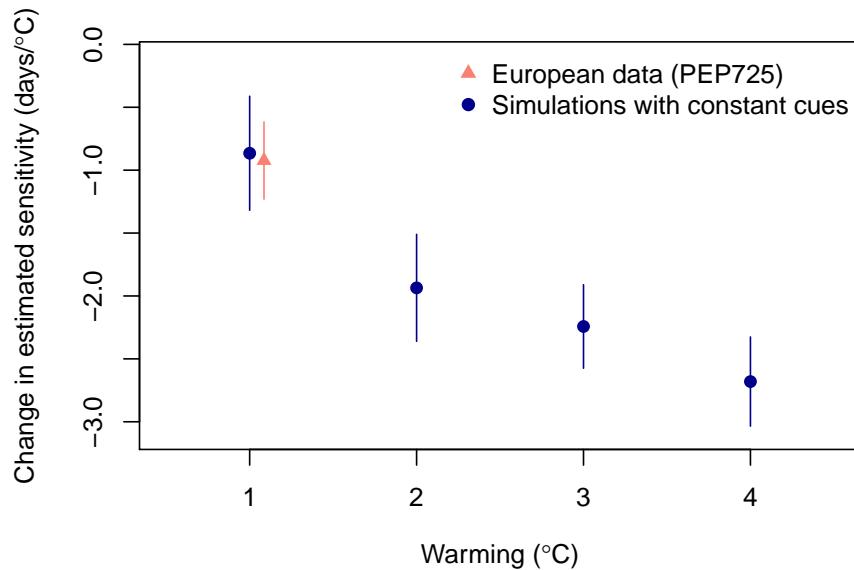


Figure S7: Declining sensitivities observed in long-term European data for a suite of common trees may be explained by a statistical artifact. We compared the sensitivity estimated from linear regressions of day of leafout versus mean spring temperature (estimated thus as days/ $^{\circ}\text{C}$ ) from PEP 725 data for *Betula pendula* from 45 sites ('European data') with estimated declines in simulations where the cues were held constant but spring temperatures warmed by 1-4 $^{\circ}\text{C}$  ('Simulations') and found the estimated temperature sensitivity measured as days/ $^{\circ}\text{C}$  declined even though the underlying cues had not changed, see *Understanding declines in temperature sensitivity in European long-term data* in Supplement for further details.

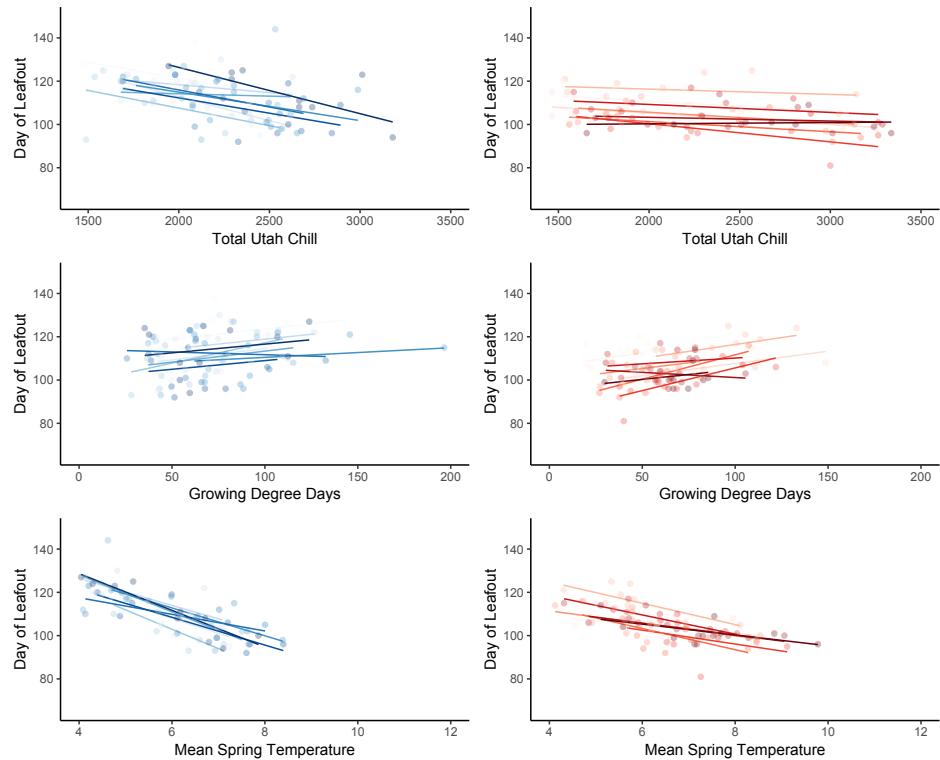
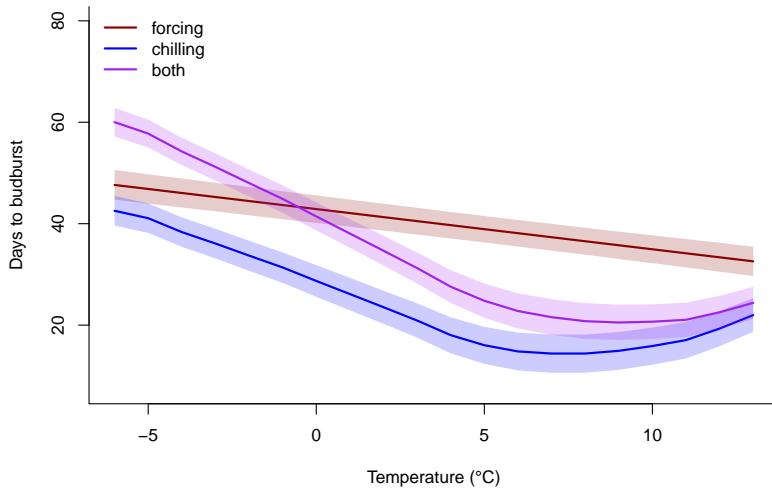
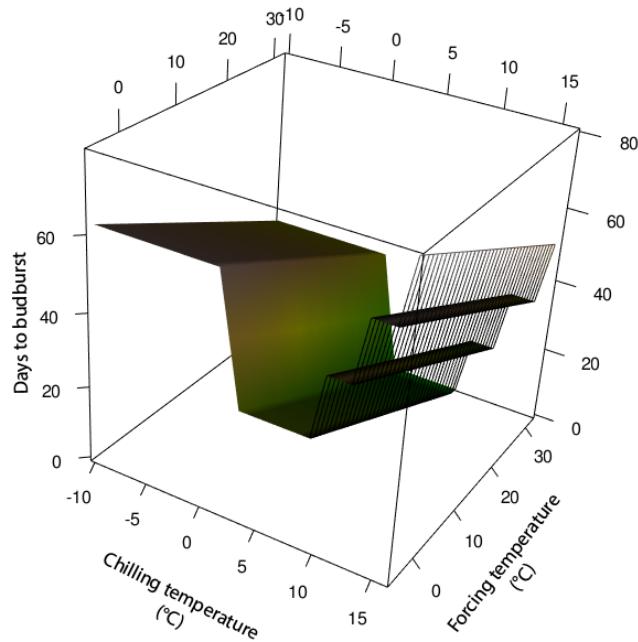


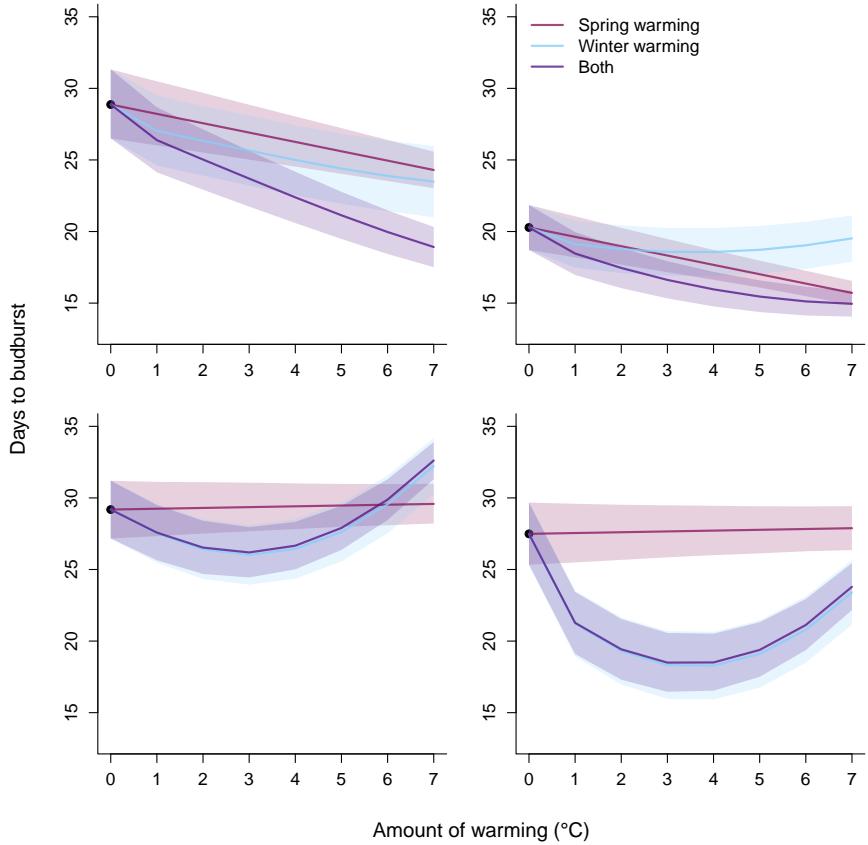
Figure S8: **Day of leaf out versus chilling, growing degree-days, and mean spring temperature** pre- (left panels, 1951-1961) and post- warming (right panels, 2000-2010) for PEP sites in Germany where *Betula pendula* phenology has been monitored for decades.



**Figure S9: Chilling and forcing temperatures affect budburst across a range of experimental conditions**, as shown here in a two-dimensional version of Fig. 2 in the main text. We predict budburst timing based on forcing temperature and estimated chilling (converted to a mean temperature, see *Estimating chilling* in the Supplemental Methods). Note that days to budburst is relative to experimental methods and thus not comparable to day of year in the field, shading represents 50 % credible intervals. We show the effect of chilling temperature on budburst, with forcing kept at the mean level across all experiments (XX °C); the effect of forcing temperature with chilling kept at the mean level across all experiments (XX chilling units), and the effect of varying both chilling and forcing simultaneously. Compare this to Figure 2 in the main text, which shows all possible combinations of chilling and forcing temperatures in a three-dimensional diagram.



**Figure S10: Based on the OSPREE model, days to budburst decrease linearly with forcing temperature and vary nonlinearly with chilling temperature due to the way that chilling is estimated (in this case, the Utah model). Forcing treatment temperatures in growth chamber experiments ranged from 0-32 °C and chilling temperatures ranged from -10-16 °C (see Table 2S for details). Budburst responses predicted by the main budburst model are shown across the full range of experimental conditions in the OSPREE database with chilling calculated as a constant temperature across a range of durations (as is commonly applied in experiments). Compare this to Figure 2 in the main text, which uses field chilling at mean chilling temperatures.**



**Figure S11: Implications of warming on budburst timing varies across species and sites,** depending strongly on pre-warming climate conditions related to chilling for each site. Here we show species-level estimates from our model (see Fig. 1 in the main text) for the two most common species in the OSPREE database: *Betula pendula* (A,B) and *Fagus sylvatica* (C,D). We compare estimates of budburst assuming varying levels of winter warming (i.e., affecting chilling, Fig. S12), with forcing kept at the mean pre-warming level, to estimates assuming varying levels of spring warming (i.e. forcing) with chilling kept at mean pre-warming levels, to estimates with winter and spring warming occurring simultaneously. Compare this to Figure 3 in the main text, which shows all possible combinations of winter and spring warming in a three-dimensional diagram.

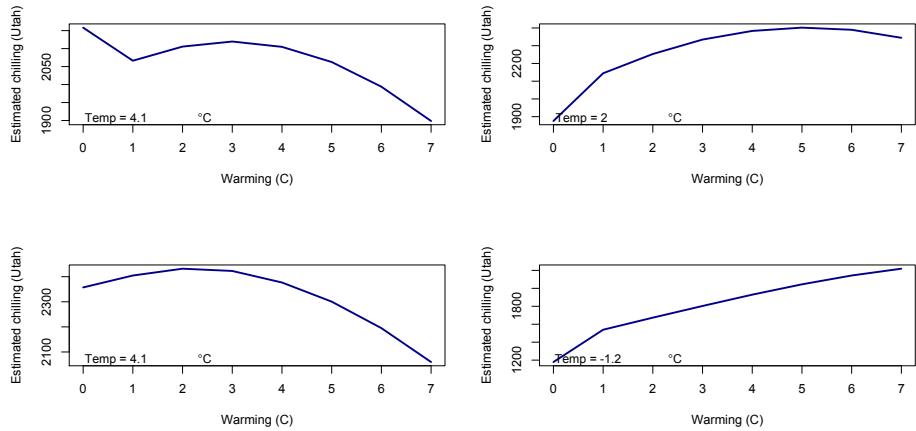


Figure S12: **Implications of global warming on chilling vary by site**, depending on pre-warming climate. For sites in A (lat, lon) and D (lat, lon), chilling increases with warming, whereas chilling decreases with warming for the sites in B (lat, lon) and C (lat, lon). Compare to Figure 3 in the main text.

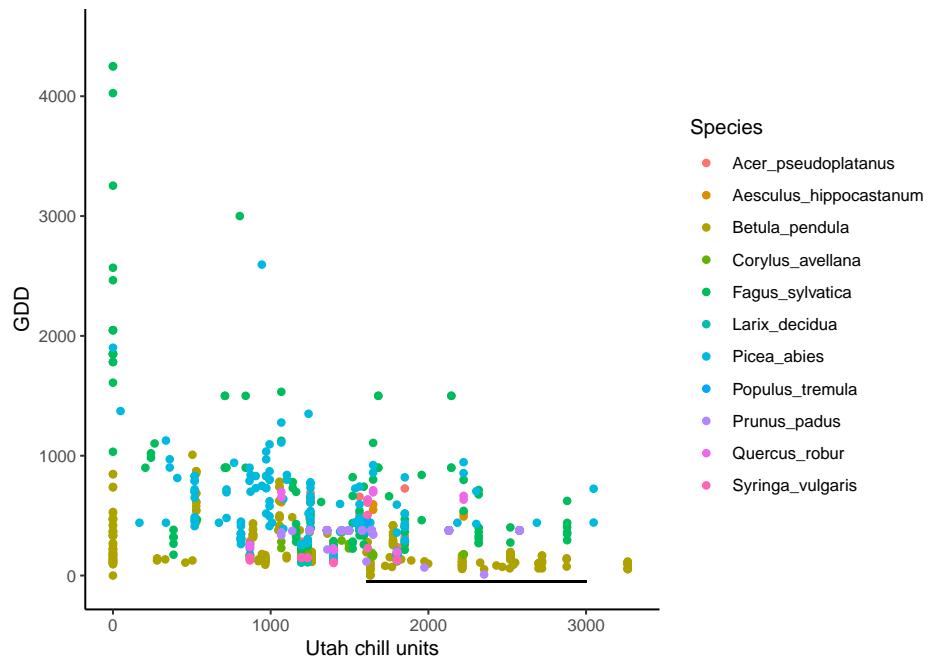
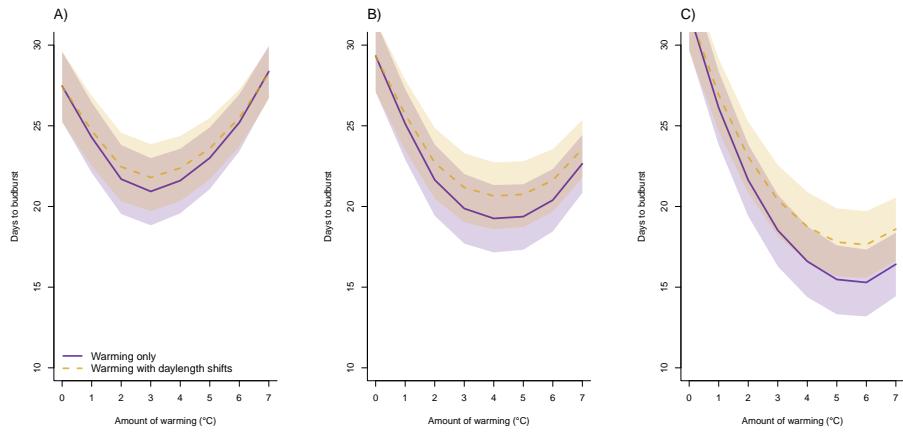


Figure S13: Growing degree days (GDD) versus chill units at the time of budburst from the OSPREE database for common species in the PEP 725 long-term phenological database. The black line shows the range of chilling (10-90% quantiles) accumulated from 1 September to 1 March for 45 sites for *Betula pendula* (see also *Understanding declines in temperature sensitivity in European long-term data*). We calculated GDD here as the average daily forcing temperature multiplied by days to budburst.



**Figure S14: Budburst is affected by climate-change induced shifts in photoperiod, especially at high latitudes,** though effects vary by site and are minor compared to effects of warming. We show forecasted effects of varying levels of warming on *Fagus sylvatica*, the most photoperiod-sensitive species in OSPREE, across three latitudes within its range, as predicted by the OSPREE model.