Methods and Supplemental Materials: Winter temperatures dominate spring phenological responses to warming

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

Methods

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

This meta-analysis follows systematic review methods (e.g., **?**) to facilitate replication and use by other researchers. Toward this end, we have created a database, which is now freely available on KNB (1) and which we hope other researchers will find to be useful. We searched the literature for research papers that 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 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. To be included, papers needed to focus on woody plants in temperate ecosystems and test for at least photoperiod or temperature effects on budburst, leafout, or flowering, and we needed to be able to quantify the phenological response to chilling, forcing, and/or photoperiod. We used ImageJ to scrape these response data from figures, whenever possible, and added additional relevant information from the tables and text of each manuscript that could not be scraped. Multiple people checked scraping and data-entry, and mis-entered data and other mistakes were cleaned. The full OSPREE database and all cleaning code is available upon request.

Although 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 72 experiments (in 49 papers, Table S1), 39 years, and 203 species (Table S2, Fig. S1).

Some species are only represented in one dataset in the OSPREE database. In these instances, it is not possible to statistically differentiate between species, study, and treatment effects. 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 congenerics in other datasets were excluded from most of our analyses, except when analyzing 'all species.'

Defining budburst

Most studies defined budburst as initial "green tips" (33/49 papers). 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 papers (4/49) did not include a definition of budburst. The majority of papers 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). OF the studiesWe 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. Although 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; 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 controlled environment conditions) and experimental chilling (*i.e.*, chilling that plant material experienced in controlled environment conditions) into two widely used metrics of chilling: Utah units and dynamic chill portions *(2)*. We used the *chillR* package (version 0.70.17) in R *(3; 4)*, version 3.6.0, to calculate both positive Utah units and dynamic chill portions from timeseries of hourly temperature data. To estimate field chilling, we generated hourly time series from a European-wide gridded climate dataset *(5)*, from which we extracted daily minimum and maximum temperature from the grid cells and dates during which experiments were conducted. For experimental chilling, we used reported chilling treatments to generate time-series of hourly temperature data.

In the formulation we used, 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 originally 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, and versions of them are routinely applied to forest trees *(e.g., 6)*. We found the effects of chilling and other cues remained qualitatively consistent across the two methods of estimating total chilling, though total chilling and photoperiod estimates were slightly lower using chill portions compared to Utah units (Table S3).

We wished to explore model predictions across a wide range of experimental temperature conditions (*i.e.*, chilling and forcing temperatures) applied by studies included in the OSPREE database (Fig. 2). To do this, we needed to convert chilling temperature to total chilling units, which could be input into our model. There is no straightforward conversion between chilling temperature and total chilling, since the duration a temperature is applied affects chilling (Fig. S5). We therefore made these conversions using two alternative approaches and we present both. For one approach, we generated daily time series of a range of experimental chilling temperatures for a range of durations spanning those in the OSPREE database (from -10 to 16°C for 7 to 240 days). We averaged across the range of durations for each temperature to get one chilling estimate per chilling temperature (Fig. S8). For the alternative approach, we used historical climate data from a gridded climate dataset *(*E-OBS*, 5)* to estimate chilling, and used these historical relationships between mean winter temperature and total chilling to convert chilling temperature to a representative amount of total chilling (Fig. 2). We present this alternative approach in the main text as it is more closely related to field chilling conditions, which was by far the most common type of chilling across experiments.

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 constant (*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 studies that ramped forcing, we calculated a weighted average of forcing temperature over the period from when forcing treatments were applied until budburst day. For studies that ramped photoperiod, we used the photoperiod individuals initially experienced (*e.g.*, studies with photoperiod lengthening from 6 hours until budburst, we recorded as '6'). When forcing and photoperiod treatments were reported as ambient, we used the E-OBS dataset to estimate mean forcing temperature and the R package *geosphere* to get daylength associated with each date and latitude *(5)*.

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 collection 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 *(7)*. This was necessary because the range and scale of each predictor varied widely (total chilling ranged from -1304 to 4724 Utah units; forcing ranged from -5.2 to 32°C, photoperiod ranged from 6 to 24 hours). We followed well-established methods of subtracting the mean and dividing by the standard deviation *(7)* to yield 'z-score' values for all predictor variables (total chilling units, forcing temperatures, and photoperiods in the experiments). In addition to these models with standardized predictors (Table S3), we also fit models in which predictors were not standardized (Table S4) so that estimates could be more easily interpreted on their natural scales. For all figures in which predictors are shown on their natural scales, we use estimates from models in which predictors were not standardized.

All models were fit using the programming languages `Stan` *(8)*(`www.mc-stan.org`), accessed via the *rstan* package (version 2.18.0) in R (3; 9), version 3.6.0. Stan provides efficient MCMC sampling via a No-U-Turn Hamiltonian Monte Carlo approach (more details can be found in *(10)* and in *(8)*). We validated our models using test data, then fit the following models described below. In all models i represents each unique observation, *sp* is the species or species complex grouping, α terms represent intercepts while β terms represent slope estimates, y is the days to budburst since forcing conditions were applied.

1. Main budburst model:

$$y_i = \alpha_{sp[i]} + \beta_{forcing_{sp[i]}} + \beta_{photoperiod_{sp[i]}} + \beta_{chilling_{sp[i]}} + \epsilon_i$$

,

$$\epsilon_i \sim N(0, \sigma_y^2)$$

The α and each of the three β coefficients were modeled at the species level, as follows:

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

We applied this model to both a dataset with 203 species ('all species'), as well as with 67 species grouped into 37 species or species complex groupings, Tables S3, S4). We present estimates from the model fit to the reduced dataset in the main text (including for Figs. 1-3 in the main text) as it represents species that were more well-represented across multiple papers and study designs, and thus are likely to be more accurate estimates (more details above in section describing the OSPREE database). Based on our modeling approach, species from fewer studies will be pooled towards the overall mean. The reduced dataset model also excluded studies which reported only 'ambient' forcing and photoperiod; these studies were included in the dataset with 203 species ('all species' model).

2. Latitude model: Given continuing debate over the role of photoperiod on budburst timing across a species' latitudinal range *(e.g., 11; 12)*, 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 collection latitude (i..e., the latitude from which plant material was collected prior to being placed in experimental conditions) 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 *(12; 13; 14)*.

   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 across different latitudes. This yielded the following model:

$$y_i = \alpha_{sp[i]} + \beta_{forcing_{sp[i]}} + \beta_{photoperiod_{sp[i]}} + \beta_{chilling_{sp[i]}} + \beta_{latitude_{sp[i]}} + \beta_{photoperiodxlatitude_{sp[i]}} + \epsilon_i,$$

$$\epsilon_i \sim N(0, \sigma_y^2)$$

The α and each of the five β coefficients were modeled at the species level, as follows:

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

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 and then exposed to 'forcing' conditions, with the assumption that tissues collected later experience more field chilling *(15)*) 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:

$$y_i = \alpha_{sp[i]} + \beta_{forcing} + \beta_{photoperiod} + \beta_{chilling} + \beta_{chillmethod} + \beta_{forcing:chillmethod} + \beta_{photoperiod:chillmethod} + \beta_{chilling:chillmethod} + \epsilon_i,$$

$$\epsilon_i \sim N(0, \sigma_y^2)$$

The α coefficients were modeled at the species level, as follows:

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

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 2 500 sampling iterations (1 500 of which were used for warm-up), yielding 4 000 posterior samples for each parameter. We assessed model performance through $\hat{R}$ 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 *(10)*.

In our figures we show means ± 50% credible intervals from our models (Figs. 1, S3 - **??**), 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 *(8; 10)*. See Tables S3- S6 for 95% credible intervals.

Modeling limitations based on experimental designs

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 (Table S8).

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 appropriate for phenological cues, especially in nature, where chilling may always be very high or extremely short photoperiods are rare. Thus we tested a non-linear (sigmoidal) model on the OSPREE data *(16)*. 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-squared = 0.53 versus 0.57), did not qualitatively alter estimates of forcing (-0.83 versus -0.79) or photoperiod (-0.25 versus -0.54) 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 are currently available.

The few studies that did incorporate interactions generally used the Weinberger method, which is not designed to robustly tease out of the effects of multiple cues (Table S6, Fig. S7). Similarly we found variation in study material/tissue and variation in thermoperiodicity were too infrequent to test for effects with current data. Our estimated effects therefore average over interactions *(7)*, but identifying them in future research will be critical to understanding and predicting budburst. This will be particularly challenging for forcing and chilling, as a lack of information on endodormancy requirements makes disentangling forcing from chilling conditions impossible with current data *(17)*.

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 ranged from 0-32°C and chilling temperatures ranged from -10-16°C in experiments, as defined by each study's authors, Figs. 2, **??**, S8). We also wished to understand how our findings may apply to conditions more commonly found in nature, where conditions often vary dramatically from those applied in controlled environment experiments. For example, very low amounts of chilling can be applied in experiments compared to the natural chilling found in many temperate areas (Fig. S5). Additionally, chilling temperature and total chilling are more correlated in nature than in experimental conditions (Fig. S5). Further, given the importance of chilling and forcing combined with the fact that seasons do not always warm uniformly with climate change *(18; 19)*, we also wished to understand how warming in the winter, spring, or both seasons would shift budburst timing. Given these goals we focused on applying our model estimates to defined levels of warming layered onto historical climate. Alternative approaches, such as using climate projections from global circulation models, would have hindered our efforts to understand degrees of warming in different seasons. Further, we emphasize that our predictions are not designed to be accurate forecasts of future budburst dates, even for the locations for which we use historical climate and budburst data. They are designed, however, to provide insights into how natural conditions can differ from experimental conditions, and to provide guidance on how much varying effects of winter and spring warming together will shape future budburst timing.

We thus applied our model to Central Europe, a well-studied area for phenology, which has both relatively long-term daily temperature data and budburst data. We selected sites that are part of the Pan European Phenology Project (PEP725, `http://www.pep725.eu`) and included data for two common European species that are prevalent in the OSPREE database: *Betula pendula* (silver birch) and *Fagus sylvatica* (European beech) *(20)*. We used a European-wide gridded climate dataset *(*E-OBS*, 5)* to extract daily minimum and maximum temperature for the grid cells where observations of leafout for these two species were available. We extracted temperature data from 1951 through 1960 (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 the R package *chillR*, see details above in *Estimating chilling* section) and mean spring forcing estimated as the mean temperature from 1 March through 30 April. We inputted these estimates for chilling and forcing into our main model, and set photoperiod to the daylength on the mean day of leafout across the PEP725 observations from from 1951 through 1960. This yielded estimates of budburst under 'pre-warming conditions,' and we then investigated model predictions of budburst given different levels of warming (from 1-7°C) above this baseline, including a full matrix of altered total chilling and forcing estimates (Figs. 3, **??**, S10).

We applied our model at all latitudes and longitudes included in the PEP725 database between 1951 and 1960 for *Betula pendula* (Fig. S6). We selected two of these sites for *Betula pendula*, as well as two sites where *Fagus sylvatica* occurs, 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 (Figs. 4, S6 - S10).

We also applied our latitude model to Central Europe, focusing on PEP725 sites where *Fagus sylvatica* leafout data were available from 1951-1960. We fit the model to three sites that differed in latitude, following the approach above for estimating baseline chilling and forcing for these sites (Fig. S11) and applying warming levels ranging from 1 to 7°C. For each site, we used as a baseline photoperiod the daylength on the mean day of leafout from PEP725 observations between 1951 and 1960. We then further estimated potential changes in photoperiod due to advancing phenology. To do this, we first estimated the shift in days to budburst as described above for Fig. 4. We then used this budburst date to estimate 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 (Fig. S11).

Note that, as described above in *Models*, our days to budburst estimate is the days to budburst since forcing conditions were applied in the experiment, which we stress is not necessarily the days to budburst after the start of ecodormancy *(17)*.

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., 21)*, 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/°C. However, if warming produces systematically higher daily temperatures this method will inherently estimate lower sensitivities, 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 PEP725 data *(20)* 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 22). 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 PEP725 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°C. We then estimated temperature sensitivity (days/°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°C with a variance of 3°C, and estimated temperature sensitivity for 10-year periods for 45 simulated sites (these values were chosen to best match the PEP725 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 PEP725 declined across the two time periods in step with warming. Across the sites studied here we estimated a decline of 0.8±0.3 days/°C (comparing 2001-2010 and 1951-1960) and 1.1±0.2°C warming; this estimate was very similar to simulations given constant cues and 1°C warming (Fig. S12).

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 warming should increase variance in leafout timing *(23)*. 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 PEP725 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 PEP725 data across the two time points (2247±31 Utah units in 1951-1960, compared to 2236±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 (Figs. S13, S14).

This potential artifact adds to existing research that has documented the statistical challenges of accurately estimating temperature sensitivities from long-term data *(24; 25)* 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 this artifact. Indeed, in the PEP725 data we found little difference across the two time-periods in GDD (68.7±2.6 in 1950-1960 versus 61.5±2.0 in 2000-2010 for GDD calculated from January 1st to leafout with a base temperature of 0°C; and a mean temperature in the 30 days before leafout of 6.8°C±0.1 in 1950-1960 versus 6.6°C±0.1). Methods such as these (that accumulate thermal temperatures until event date) are also vulnerable to other issues: as researchers must select the day to start accumulating or averaging temperatures, these methods should work best when the start date is always after endodormancy break, when plants are most responsive to forcing *(17)*. 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 *(24)*. Without measures of endodormancy break *(17)*, we suggest efforts to accurately estimate cues from long-term observational data may be difficult or impossible without additional physiological information from controlled environment experiments.

Applying the sliding windows approach to Central European data

We also tested whether the observed declines in sensitivity could be explained via the absolute sliding time window (SWA) approach (26). CAT: please add some description here.

We found that this approach yielded consistent results with our previous analyses: higher sensitivity during the pre-warming time period (Table XS).

Data Availibility

The OSPREE database is available at Knowledge Network for Biocomplexity, doi:10.5063/F1QV3JQR *(1)*.

Code Availibility

All code is available on GitHub at https://github.com/lizzieinvancouver/ospree/tree/bbculdesac.

References

[1] E. M. Wolkovich, *et al.*, Observed Spring Phenology Responses in Experimental Environments (OSPREE), doi:10.5063/F1QV3JQR (2019).

[2] F. Dennis, *HortScience* **38**, 347 (2003).

[3] R Development Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria (2017).

[4] E. Luedeling, *chillR: Statistical Methods for Phenology Analysis in Temperate Fruit Trees* (2019). R package version 0.70.17.

[5] R. C. Cornes, G. van der Schrier, E. J. van den Besselaar, P. D. Jones, *Journal of Geophysical Research: Atmospheres* **123**, 9391 (2018).

[6] C. A. Harrington, P. J. Gould, J. B. St Clair, *Forest Ecology and Management* **259**, 798 (2010).

[7] A. Gelman, J. Hill, *Data analysis using regression and multilevel/hierarchical models* (Cambridge University Press, 2006).

[8] B. Carpenter, *et al.*, *Journal of Statistical Software* **76**, 10.18637/jss.v076.i01 (2017).

[9] Stan Development Team, RStan: the R interface to Stan (2018). R package version 2.17.3.

[10] A. Gelman, *et al.*, *Bayesian Data Analysis* (CRC Press, New York, 2014), third edn.

[11] C. M. Zohner, B. M. Benito, J. C. Svenning, S. S. Renner, *Nature Climate Change* **6**, 1120 (2016).

[12] J. Gauzere, *et al.*, *Agricultural and Forest Meteorology* **244**, 9 (2017).

[13] K. Saikkonen, *et al.*, *Nature Climate Change* **2**, 239 (2012).

[14] D. A. Way, R. A. Montgomery, *Plant, Cell & Environment* **38**, 1725 (2015).

[15] J. H. Weinberger, *et al.*, *Proceedings. American Society for Horticultural Science* (1950), vol. 56, pp. 122–28.

[16] I. Chuine, I. Garcia de Cortazar Atauri, H. Hanninen, K. Kramer, *Plant development models* (Kluwer, Dordrecht, the Netherlands, 2013), pp. 275–293.

[17] I. Chuine, *et al.*, *Global Change Biology* **22**, 3444 (2016).

[18] R. Vautard, *et al.*, *Environmental Research Letters* **9**, 034006 (2014).

[19] D. Agency, C. Körner, Global and European temperature (2019).

[20] B. Templ, *et al.*, *International Journal of Biometeorology* **62**, 1109 (2018).

[21] Y. S. H. Fu, *et al.*, *Nature* **526**, 104 (2015).

[22] IPCC, *Climate Change 2014: Impacts, Adaptation, and Vulnerability* (Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2014).

[23] K. R. Ford, C. A. Harrington, S. Bansal, J. Gould, Peter, J. B. St. Clair, *Global Change Biology* **22**, 3712 (2016).

[24] S. Güsewell, R. Furrer, R. Gehrig, B. Pietragalla, *Global Change Biology* **23**, 5189 (2017).

[25] J. S. Clark, J. Melillo, J. Mohan, C. Salk, *Global Change Biology* **20**, 1136 (2014).

[26] E. G. Simmonds, E. F. Cole, B. C. Sheldon, *Journal of Animal Ecology* (2019).

[27] D. Basler, C. Körner, *Agricultural and Forest Meteorology* **165**, 73 (2012).

[28] D. Basler, C. Körner, *Tree Physiology* **34**, 377 (2014).

[29] L. Biasi, F. Zanette, R. Carvalho, *XXVIII International Horticultural Congress on Science and Horticulture for People (IHC2010): International Symposium on Plant 932* (2012), pp. 507–512.

[30] A. Caffarra, A. Donnelly, *International Journal of Biometeorology* **55**, 711 (2011).

[31] A. Caffarra, A. Donnelly, I. Chuine, M. B. Jones, *Climate Research* **46**, 147 (2011).

[32] S. Calmé, F. J. Bigras, H. A. Margolis, C. Hébert, *Tree Physiology* **14**, 1313 (1994).

[33] R. K. Campbell, A. I. Sugano, *Botanical Gazette* pp. 290–298 (1975).

[34] G. Chavarria, *et al.*, *Ciência Rural* **39**, 2016 (2009).

[35] C. Cook, G. Jacobs, *The Journal of Horticultural Science and Biotechnology* **75**, 233 (2000).

[36] M. Falusi, R. Calamassi, *Plant Biosystems-An International Journal Dealing with all Aspects of Plant Biology* **137**, 185 (2003).

[37] M. Falusi, R. Calamassi, *Tree Physiology* **6**, 429 (1990).

[38] M. Falusi, R. Calamassi, *Annales des Sciences forestières* (EDP Sciences, 1996), vol. 53, pp. 967–979.

[39] M. Falusi, R. Calamassi, *Plant Biosystems-An International Journal Dealing with all Aspects of Plant Biology* **131**, 143 (1997).

[40] L. Ghelardini, A. Santini, S. Black-Samuelsson, T. Myking, M. Falusi, *Tree physiology* **30**, 264 (2010).

[41] T. Gianfagna, S. Mehlenbacher, *HortScience* **20**, 909 (1985).

[42] D. Gömöry, E. Foffová, R. Longauer, D. Krajmerová, *European Journal of Forest Research* **134**, 89 (2015).

[43] S. Guak, D. M. Olsyzk, L. H. Fuchigami, D. T. Tingey, *Tree Physiology* **18**, 671 (1998).

[44] P. Guerriero, G. Scalabrelli, G. Grazzini, *I International Symposium on Kiwifruit 282* (1990), pp. 79–86.

[45] O. M. Heide, A. Sønsteby, *Scientia Horticulturae* **138**, 64 (2012).

[46] O. Heide, *Physiologia Plantarum* **88**, 531 (1993).

[47] O. Heide, *Physiologia Plantarum* **89**, 187 (1993).

[48] H. Jones, R. Hillis, S. Gordon, R. Brennan, *Plant Biology* **15**, 18 (2012).

[49] J. Laube, *et al.*, *Global Change Biology* **20**, 170 (2014).

[50] J. Laube, T. H. Sparks, N. Estrella, A. Menzel, *New Phytologist* **202**, 350 (2014).

[51] C. Li, *et al.*, *Tree physiology* **25**, 1563 (2005).

[52] T. Linkosalo, M. J. Lechowicz, *Tree Physiology* **26**, 1249 (2006).

[53] R. Man, P. Lu, *Canadian Journal of Forest Research* **40**, 1815 (2010).

[54] X. Morin, J. Roy, L. Sonié, I. Chuine, *New Phytologist* **186**, 900 (2010).

[55] T. Myking, O. Heide, *Tree physiology* **15**, 697 (1995).

[56] T. Myking, *Trees* **12**, 107 (1997).

[57] T. Myking, *Trees* **12**, 224 (1998).

[58] M. Pagter, U. B. Andersen, L. Andersen, *AoB plants* **7**, plv024 (2015).

[59] J. Partanen, I. Leinonen, T. Repo, *Silva Fennica* **35**, 111 (2001).

[60] J. Partanen, V. Koski, H. Hänninen, *Tree Physiology* **18**, 811 (1998).

[61] A. Ramos, L. Rallo, *III International Symposium on Olive Growing 474*, I. T. Metzidakis, D. G. Voyiatzis, eds. (1999), pp. 251–254.

[62] P. Rinne, A. Saarelainen, O. Junttila, *Physiologia Plantarum* **90**, 451 (1994).

[63] P. Rinne, H. Hänninen, P. Kaikuranta, J. Jalonen, T. Repo, *Plant, Cell & Environment* **20**, 1199 (1997).

[64] V. Sanz-Perez, P. Castro-Diez, F. Valladares, *Plant Biology* **11**, 142 (2009).

[65] V. Sanz-Pérez, P. Castro-Díez, *Trees* **24**, 89 (2010).

[66] B. J. Schnabel, R. L. Wample, *American Journal of Enology and Viticulture* **38**, 265 (1987).

[67] R. Skuterud, J. Dietrichson, *Silva Fennica* **28**, 223 (1994).

[68] A. Sønsteby, O. M. Heide, *Scientia Horticulturae* **179**, 256 (2014).

[69] T. M. Spann, J. G. Williamson, R. L. Darnell, *Journal of the American Society for Horticultural Science* **129**, 294 (2004).

[70] J. Spiers, A. Draper, *Journal American Society for Horticultural Science* (1974).

[71] H. Swartz, L. Powell Jr, *Symposium on Growth Regulators in Fruit Production 120* (1981), pp. 173–178.

[72] B. Thielges, R. Beck, *Tree Physiology and Yield Improvement* **14**, 253 (1976).

[73] D. P. Webb, *Forest Science* **23**, 474 (1977).

[74] J. Worrall, F. Mergen, *Physiologia Plantarum* **20**, 733 (1967).

Supplemental Tables

Table S1: **Dataset names and references for papers in the OSPREE database.**

Dataset	Reference
basler12	(27)
basler14	(28)
biasi12	(29)
caffarra11a	(30)
caffarra11b	(31)
calme94	(32)
campbell75	(33)
chavarria09	(34)
cook00b	(35)
falusi03	(36)
falusi90	(37)
falusi96	(38)
falusi97	(39)
ghelardini10	(40)
gianfagna85	(41)
gomory15	(42)
guak98	(43)
guerriero90	(44)
heide12	(45)
heide93	(46)
heide93a	(47)
jones12	(48)
laube14a	(49)
laube14b	(50)
li05	(51)
linkosalo06	(52)
man10	(53)
morin10	(54)
myking95	(55)
myking97	(56)
myking98	(57)
pagter15	(58)
partanen01	(59)
partanen98	(60)
ramos99	(61)
rinne94	(62)
rinne97	(63)
Sanz-Perez09	(64)
sanzperez10	(65)
schnabel87	(66)
skuterud94	(67)
sonsteby14	(68)
spann04	(69)
spiers74	(70)
swartz81	(71)
thielges75	(72)
webb78	(73)
worrall67	(74)
zohner16	(11)

Table S2: **Species included in the OSPREE database.** See Table S1 for reference associated with each dataset.

Species	Number of papers	Dataset
Abies alba	2	basler12, laube14a
Abies homolepis	1	laube14a
Acer barbinerve	1	zohner16

Table S2: **Species included in the OSPREE database.** See Table S1 for reference associated with each dataset.

Species	Number of papers	Dataset
Acer campestre	1	zohner16
Acer ginnala	1	zohner16
Acer negundo	1	laube14a
Acer platanoides	1	zohner16
Acer pseudoplatanus	3	basler12, basler14, laube14a
Acer saccharinum	1	webb78
Acer saccharum	3	calme94, laube14a, webb78
Acer tataricum	1	laube14a
Actinidia deliciosa	2	biasi12, guerriero90
Aesculus flava	1	zohner16
Aesculus hippocastanum	3	basler12, laube14a, zohner16
Aesculus parviflora	1	zohner16
Alnus glutinosa	2	heide93, myking98
Alnus incana	2	heide93, zohner16
Alnus maximowiczii	1	zohner16
Amelanchier alnifolia	1	zohner16
Amelanchier florida	1	zohner16
Amelanchier laevis	1	zohner16
Amorpha fruticosa	1	laube14a
Aronia melanocarpa	1	zohner16
Berberis dielsiana	1	zohner16
Betula alleghaniensis	1	calme94
Betula lenta	1	zohner16
Betula nana	1	zohner16
Betula pendula	10	heide93, li05, rinne97, basler12, laube14a, laube14b, linkosalo06, myking95, skuterud94
Betula populifolia	1	zohner16
Betula pubescens	6	heide93, rinne94, caffarra11a, caffarra11b, myking95, myking97
Buddleja albiflora	1	zohner16
Buddleja alternifolia	1	zohner16
Buddleja davidii	1	zohner16
Caragana pygmaea	1	zohner16
Carpinus betulus	3	heide93a, laube14a, zohner16
Carpinus laxiflora	1	zohner16
Carpinus monbeigiana	1	zohner16
Carya cordiformis	1	zohner16
Carya laciniosa	1	zohner16
Carya ovata	1	zohner16
Castanea sativa	1	zohner16
Cedrus libani	1	zohner16
Celtis caucasica	1	zohner16
Celtis laevigata	1	zohner16
Celtis occidentalis	1	zohner16
Cephalanthus occidentalis	1	zohner16
Cercidiphyllum japonicum	1	zohner16
Cercidiphyllum magnificum	1	zohner16
Cercis canadensis	1	zohner16
Cercis chinensis	1	zohner16
Cladrastis lutea	1	zohner16
Cornus alba	2	laube14a, zohner16
Cornus kousa	1	zohner16
Cornus mas	2	laube14a, laube14b
Corylopsis sinensis	1	zohner16
Corylopsis spicata	1	zohner16
Corylus avellana	4	basler12, heide93, laube14a, zohner16
Corylus heterophylla	1	zohner16
Corylus sieboldiana	1	zohner16

Table S2: **Species included in the OSPREE database.** See Table S1 for reference associated with each dataset.

Species	Number of papers	Dataset
Decaisnea fargesii	1	zohner16
Deutzia gracilis	1	zohner16
Deutzia scabra	1	zohner16
Elaeagnus ebbingei	1	zohner16
Eleutherococcus senticosus	1	zohner16
Eleutherococcus setchuenensis	1	zohner16
Eleutherococcus sieboldianus	1	zohner16
Euonymus europaeus	1	zohner16
Euonymus latifolius	1	zohner16
Fagus crenata	1	zohner16
Fagus engleriana	1	zohner16
Fagus orientalis	1	zohner16
Fagus sylvatica	10	falusi03, falusi90, falusi96, falusi97, basler12, basler14, caffarra11a, heide93a, zohner16
Forsythia ovata	1	zohner16
Forsythia suspensa	1	zohner16
Fraxinus americana	1	webb78
Fraxinus chinensis	1	laube14a
Fraxinus excelsior	2	basler12, laube14a
Fraxinus latifolia	1	zohner16
Fraxinus ornus	1	zohner16
Fraxinus pennsylvanica	1	laube14a
Ginkgo biloba	1	zohner16
Hamamelis japonica	1	zohner16
Hamamelis vernalis	1	zohner16
Heptacodium miconioides	1	zohner16
Hibiscus syriacus	1	zohner16
Hydrangea arborescens	1	zohner16
Hydrangea involucrata	1	zohner16
Hydrangea serrata	1	zohner16
Juglans ailantifolia	1	laube14a
Juglans cinerea	1	laube14a
Juglans regia	1	laube14a
Larix decidua	4	basler12, gomory15, laube14a, laube14b
Larix gmelinii	1	zohner16
Larix kaempferi	1	zohner16
Ligustrum tschonoskii	1	zohner16
Liquidambar orientalis	1	zohner16
Liquidambar styraciflua	1	zohner16
Liriodendron tulipifera	1	zohner16
Lonicera alpigena	1	zohner16
Lonicera caerulea	1	zohner16
Lonicera maximowiczii	1	zohner16
Malus domestica	3	cook00b, gianfagna85, swartz81
Metasequoia glyptostroboides	1	zohner16
Nothofagus antarctica	1	zohner16
Oemleria cerasiformis	1	zohner16
Olea europaea	1	ramos99
Orixa japonica	1	zohner16
Ostrya carpinifolia	1	zohner16
Ostrya virginiana	1	zohner16
Paeonia rockii	1	zohner16
Parrotia persica	1	zohner16
Parrotiopsis jaquemontiana	1	zohner16
Photinia villosa	1	zohner16

Table S2: **Species included in the OSPREE database.** See Table S1 for reference associated with each dataset.

Species	Number of papers	Dataset
Picea abies	9	basler12, basler14, gomory15, laube14a, laube14b, partanen01, partanen98, worrall67
Picea glauca	1	man10
Pinus nigra	1	laube14a
Pinus strobus	1	laube14a
Pinus sylvestris	1	laube14a
Pinus wallichiana	1	laube14a
Populus deltoides	1	thielges75
Populus koreana	1	zohner16
Populus tremula	3	heide93, laube14a, laube14b
Prinsepia sinensis	1	zohner16
Prinsepia uniflora	1	zohner16
Prunus avium	2	basler12, laube14a
Prunus cerasifera	1	zohner16
Prunus padus	3	heide93, myking98, zohner16
Prunus persica	1	chavarria09
Prunus serotina	1	laube14a
Prunus serrulata	1	zohner16
Prunus tenella	1	zohner16
Pseudotsuga menziesii	3	guak98, campbell75, laube14a
Ptelea trifoliata	1	zohner16
Pyrus elaeagnifolia	1	zohner16
Pyrus pyrifolia	1	zohner16
Pyrus ussuriensis	1	zohner16
Quercus bicolor	1	laube14a
Quercus coccifera	1	sanzperez10
Quercus faginea	2	Sanz-Perez09, sanzperez10
Quercus ilex	3	Sanz-Perez09, sanzperez10, morin10
Quercus petraea	2	basler12, basler14
Quercus pubescens	1	morin10
Quercus robur	4	laube14a, laube14b, morin10, zohner16
Quercus rubra	2	calme94, laube14a
Quercus shumardii	1	zohner16
Rhamnus alpina	1	zohner16
Rhamnus cathartica	1	zohner16
Rhododendron canadense	1	zohner16
Rhododendron dauricum	1	zohner16
Rhododendron mucronulatum	1	zohner16
Ribes alpinum	1	zohner16
Ribes divaricatum	1	zohner16
Ribes glaciale	1	zohner16
Ribes nigrum	4	jones12, heide12, pagter15, sonsteby14
Robinia pseudoacacia	2	laube14a, laube14b
Rosa hugonis	1	zohner16
Rosa majalis	1	zohner16
Rubus idaeus	1	heide93
Salix gracilistyla	1	zohner16
Salix repens	1	zohner16
Salix smithiana	1	caffarra11a
Sambucus nigra	1	zohner16
Sambucus pubens	1	zohner16
Sambucus tigranii	1	zohner16
Sinowilsonia henryi	1	zohner16
Sorbus aria	1	zohner16
Sorbus aucuparia	2	basler12, heide93
Sorbus commixta	1	zohner16
Sorbus decora	1	zohner16

Table S2: **Species included in the OSPREE database.** See Table S1 for reference associated with each dataset.

Species	Number of papers	Dataset
Spiraea canescens	1	zohner16
Spiraea chamaedryfolia	1	zohner16
Spiraea japonica	1	zohner16
Stachyurus praecox	1	zohner16
Stachyurus sinensis	1	zohner16
Symphoricarpos albus	2	laube14a, laube14b
Syringa josikaea	1	zohner16
Syringa reticulata	1	zohner16
Syringa villosa	1	zohner16
Syringa vulgaris	3	basler12, laube14a, laube14b
Tilia cordata	2	basler12, caffarra11a
Tilia dasystyla	1	zohner16
Tilia japonica	1	zohner16
Tilia platyphyllos	1	zohner16
Toona sinensis	1	zohner16
Ulmus americana	1	zohner16
Ulmus glabra	1	ghelardini10
Ulmus laevis	1	zohner16
Ulmus macrocarpa	1	ghelardini10
Ulmus minor	1	ghelardini10
Ulmus parvifolia	1	ghelardini10
Ulmus pumila	1	ghelardini10
Ulmus villosa	1	ghelardini10
Vaccinium ashei	1	spiers74
Vaccinium corymbosum	1	spann04
Viburnum betulifolium	1	zohner16
Viburnum buddleifolium	1	zohner16
Viburnum carlesii	1	zohner16
Viburnum opulus	1	zohner16
Viburnum plicatum	1	zohner16
Vitis vinifera	2	biasi12, schnabel87
Weigela coraeensis	1	zohner16
Weigela florida	1	zohner16
Weigela maximowiczii	1	zohner16

Table S3: **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 to the Utah model. These models were fit to a dataset including species that are well-represented in the OSPREE database, with 37 species or species complexes (67 unique species). We also present coefficients from a model including all species with all treatment types (with no species complexes used). We present posterior means, as well as 50 percent and 95 percent credible intervals from models in which the predictors have been standardized so that they are directly comparable.

	Utah units					chill portions					All species, Utah units				
	mean	25%	75%	2.5%	97.5%	mean	25%	75%	2.5%	97.5%	mean	25%	75%	2.5%	97.5%
μ_{α}	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
σ_{α}	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
σ_{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 S4: **Estimates from models fit with predictors on their natural scales**, so that estimates 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 and here we also present estimates from a model with chilling in chill portions, with both fit to a dataset including species that are well-represented in the OSPREE database, with 37 species or species complexes (67 unique species). We also present coefficients from a model including all species and all treatment types (with no species complexes used). We present posterior means, as well as 50 precent and 95 percent credible intervals, from models.

	Utah units					chill portions					All species, Utah units				
	mean	25%	75%	2.5%	97.5%	mean	25%	75%	2.5%	97.5%	mean	25%	75%	2.5%	97.5%
μ_{α}	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
σ_{α}	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
σ_{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 S5: **Estimates from latitude model fit with standardized predictors**. Using a model with Utah chilling units and testing the effects of collection 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 50 percent and 95 percent credible intervals from models in which the predictors have been standardized so that they are directly comparable,

	mean	25%	75%	2.5%	97.5%
μ_{α}	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
σ_{α}	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
σ_{y}	15.44	15.25	15.63	14.90	16.01
N_{sp}	36				

Table S6: **Estimates from chilling study design 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, we 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 50 percent and 95 percent credible intervals from models in which the predictors have been standardized so that they are directly comparable.

	mean	25%	75%	2.5%	97.5%
μ_{α}	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
σ_{α}	13.34	11.28	14.96	8.81	20.24
σ_{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				

Table S7: **Estimates from the life stage model fit with standardized predictors**. Using a model with Utah chilling units and testing the effects of life stage (juveninle versus adult) 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 50 percent and 95 percent credible intervals from models in which the predictors have been standardized so that they are directly comparable.

	mean	25%	75%	2.5%	97.5%
μ_{α}	29.65	28.58	30.75	26.31	32.89
$\beta_{forcing}$	-4.53	-5.31	-3.75	-6.82	-2.28
$\beta_{photoperiod}$	-3.37	-4.16	-2.58	-5.79	-1.04
$\beta_{chilling}$	-8.07	-9.09	-7.08	-11.22	-5
β_{stage}	3.56	2.69	4.41	1.02	6.11
σ_{y}	15.78	15.58	15.97	15.25	16.33
N_{sp}	37				

Table S8: **Number of studies testing for interactions between chilling, forcing, and photoperiod treatments**, out of the 42 studies (across 30 papers) included in the main budburst model

Treatment.1	Treatment.2	n.studies
photo	force	5
chilltemp	force	1
chilldays	force	5
chilltemp	photo	1
chilldays	photo	7
fieldsample.date	force	7
fieldsample.date	photo	10

Table S9: **Locations and pre-warming winter and spring conditions** for sites included in Figs. 3, **??**, S10. Units are degrees for latitude and longitude, and °C for temperature.

Species	Latitude	Longitude	Spring Temperature	Winter Temperature
Betula pendula	46.82	12.80	0.69	-1.05
	48.32	15.82	7.13	4.31
Fagus sylvatica	48.78	15.40	4.76	2.25
	46.72	15.77	7.40	4.33

Supplemental Figures

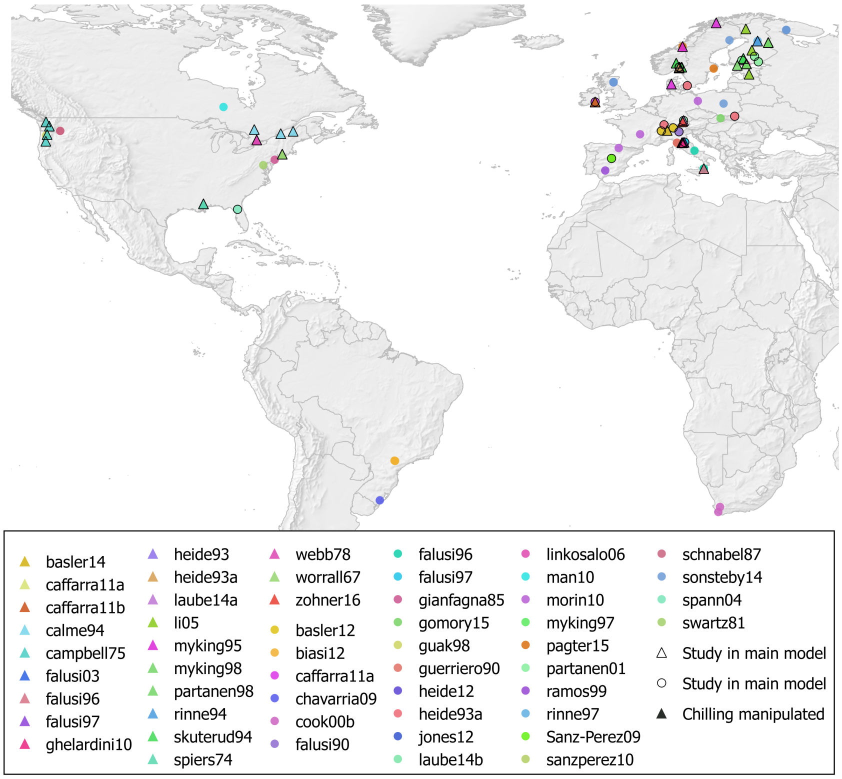

Figure S1: **Map of days to budburst experiments in the OSPREE database.** Legend shows each dataset included in the main OSPREE model with all species and treatments (Tables S3, S4); symbols outlined in black represent datasets for which chilling was manipulated experimentally or through multiple field sample dates. See Table S1 for the reference associated with each dataset.

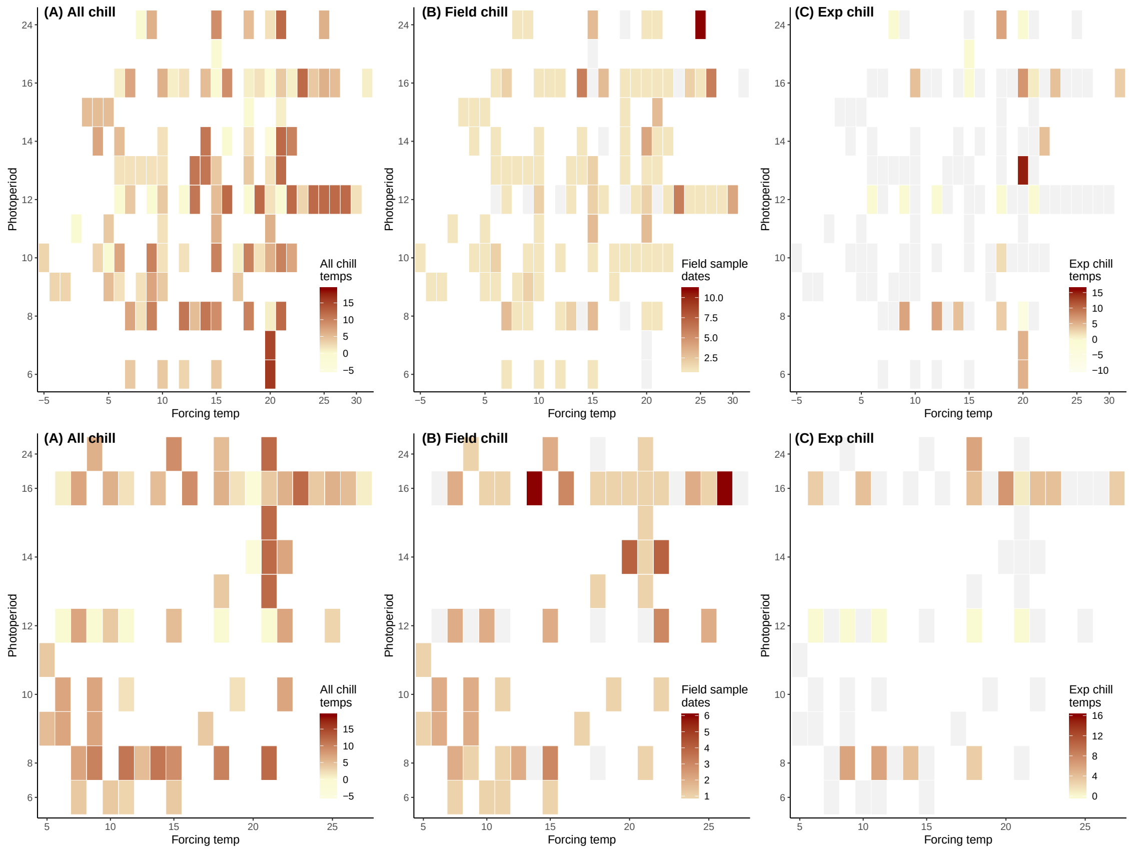

Figure S2: **The diversity of study designs used in analyses.** Heatmaps show the range and commonness of different forcing (x-axis in all panels) by photoperiod (y-axis in all panels) combinations and with which chilling they were combined. In (A, top and bottom) we show our estimated chill units, which integrate across field (B, top and bottom) and experimental chilling (C, top and bottom). The top row shows all data included in the full model with 203 species, while the bottom row shows the data included in the main model with a subset of species well-represented across treatments and studies. Gray squares indicate a treatment was not applied (*i.e.*, the prevalence of gray squares in (C) highlights how few studies include experimental chilling). Field sample dates are counted as any reported sampling dates more then 14 days apart.

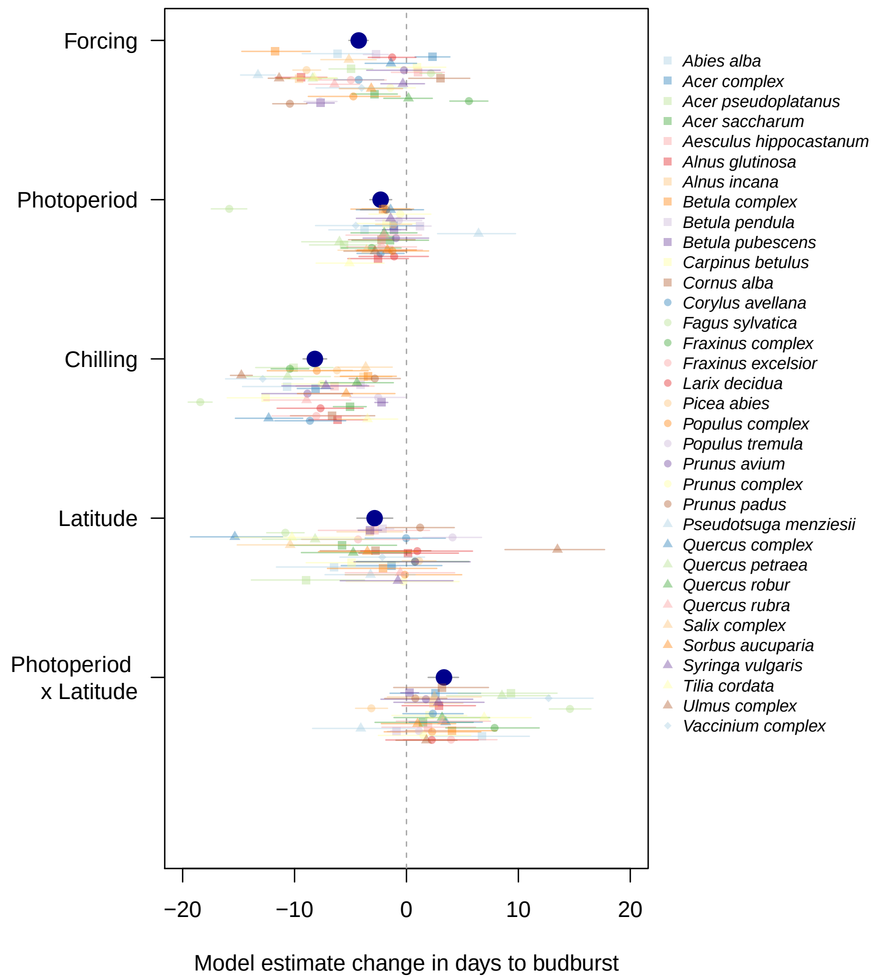

Figure S3: **Estimates for effects of chilling exceeded estimates for forcing, photoperiod, collection latitude, and the interaction between latitude and photoperiod, for most species,** in the latitude budburst model, using Utah units (Table S5). Using standardized units, which allow comparisons across cues, we show that, as with the main budburst model (Fig. 1), most species (smaller symbols) are responsive to most cues. Chilling is the strongest cue when considering overall estimates across species (larger, dark blue circles).

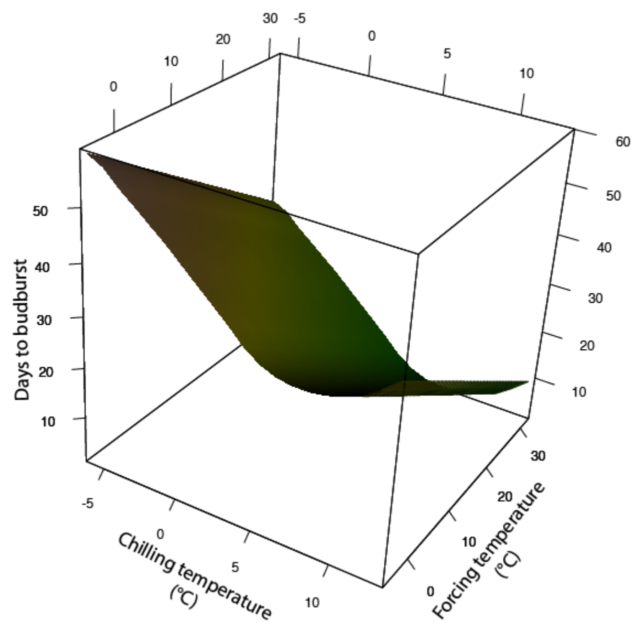

Figure S4: **Estimates of budburst across a range of forcing temperatures and estimated chilling** (converted to a representative mean temperature, see *Estimating chilling* in the Supplemental Materials) based on overall estimates of chilling and forcing effects (Fig. 1). Maximum advances in budburst occur at intermediate chilling temperatures (*e.g.*, here at mean winter temperatures of 6-7°C) and the highest forcing (here at 32°C). We set photoperiod to eight hours, which is the most common photoperiod treatment in the database. Note that days to budburst is relative to experimental methods and thus is not comparable to day of year in the field; shading represents days to budburst. Compare this to Fig. 2 in the main text.

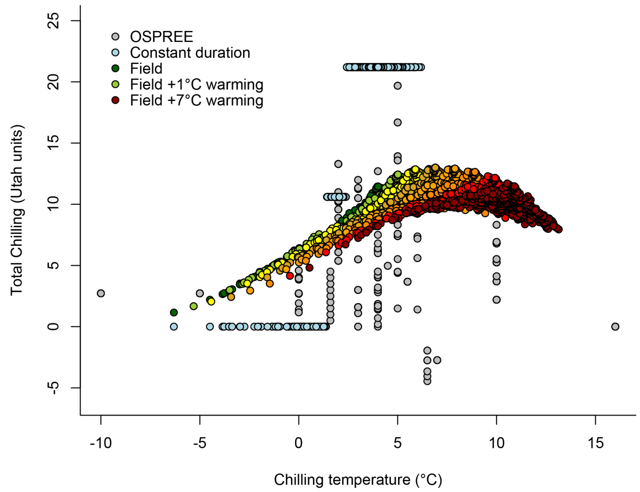

Figure S5: **Chilling accumulates differently in experiments with constant temperatures versus natural systems**, in which temperature is more strongly correlated with chilling. See *Estimating chilling* for a detailed description of 'Field' climate data, for which we use historical climate data from Europe. Fig. 2 and **??** use 'Field' relationships to convert chill temperature to total chilling, whereas Fig. S8 uses 'Constant duration' relationships.

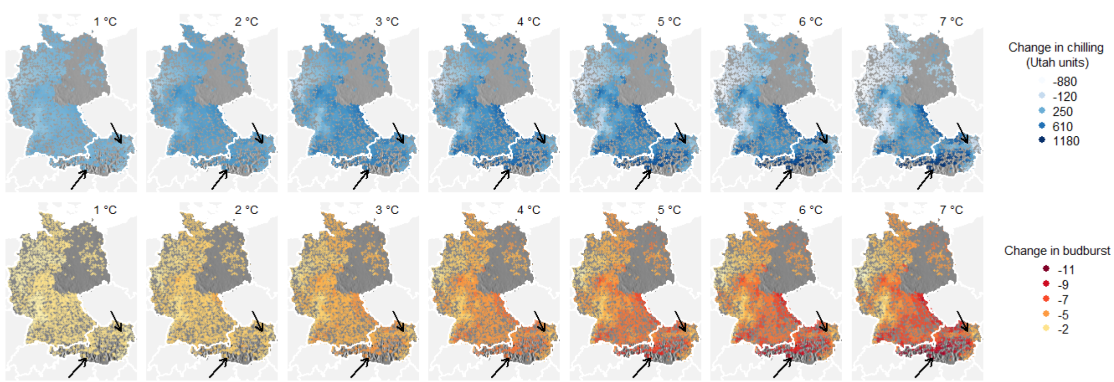

Figure S6: **Forecasted changes in chilling (top panel) and spring phenology for *Betula pendula* (bottom panel)**, in locations included in the PEP725 database, where phenology dates are known for the pre-warming time period (1951-1960). Changes in chilling and budburst are calculated relative to the mean chilling and leafout dates during this pre-warming time period for each location. Arrows indicate sites shown in Fig. 3A (latitude = 46.8°N, longitude = 12.8°E, 659 m above sea level) and 3B (latitude = 48.3°N, longitude = 15.8°E, 210 m above sea level) in the main manuscript.

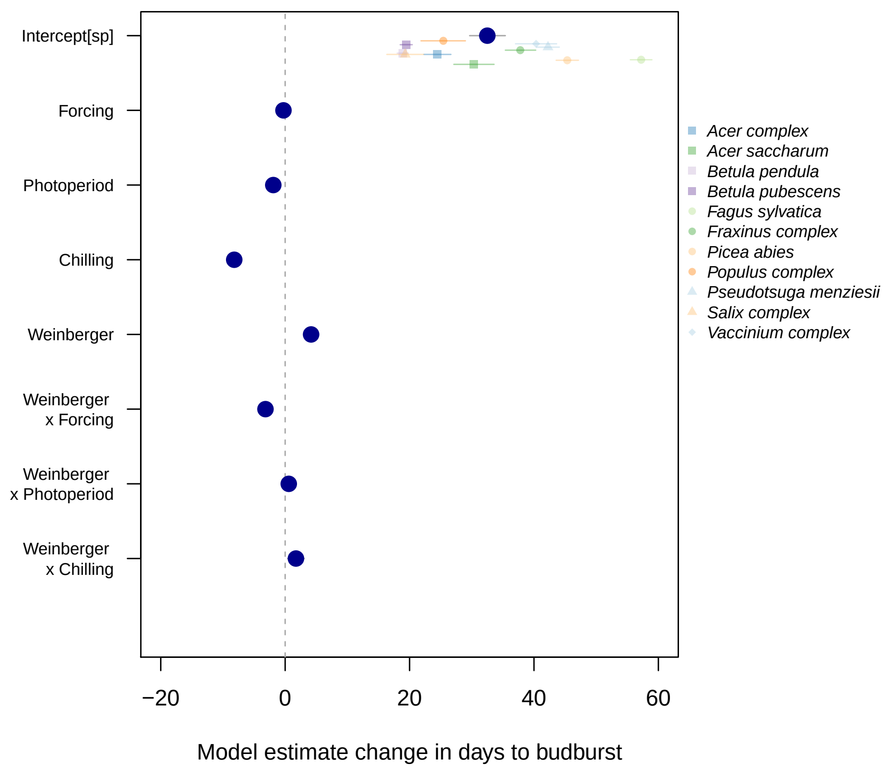

Figure S7: **Chilling study design affects estimates of major cues.** Studies using the 'Weinberger' method (sequential removal of tissue from field) had later budburst timing, stronger estimates of forcing (Weinberger x Forcing) and weaker estimates of chilling (Weinberger x Chilling) compared to studies that experimentally manipulated chilling directly. For an extended description of model and underlying data see *Chilling study design model*; for model summary see Table S6.

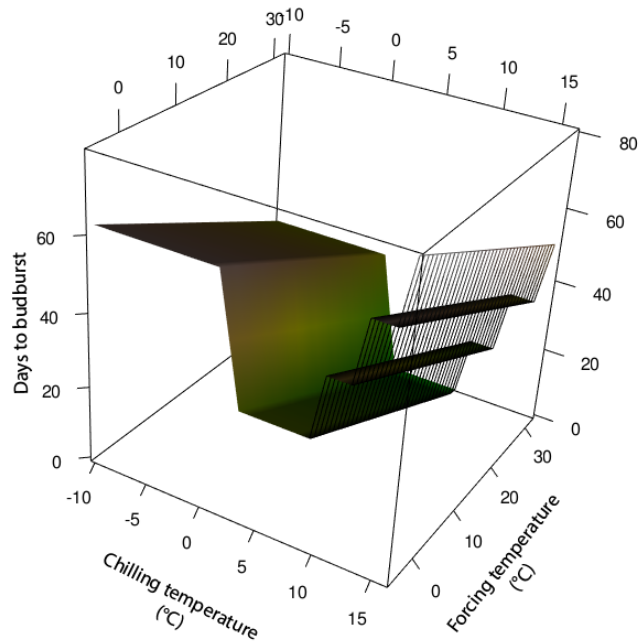

Figure S8: **Based on our model (see 'main budburst model'** ***Models*** **section), days to budburst decrease linearly with forcing temperature and vary nonlinearly with chilling temperature** due to the way that chilling is estimated (in this case, the Utah model). Forcing treatment temperatures in 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). Maximum advances in budburst occur at intermediate chilling temperatures (*e.g.*, here at mean chilling temperatures between 3 and 9°C) and the highest forcing temperatures (32°C). Compare this to Fig. 2 in the main text, which is a simplified two-dimensional version.

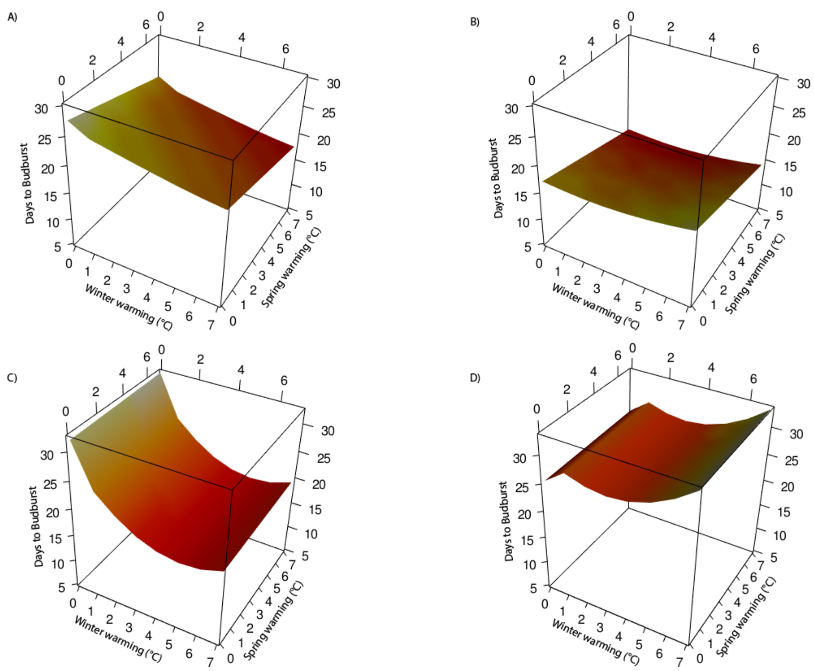

Figure S9: **Implications of warming on budburst timing varies across species and sites**, depending strongly on pre-warming climate conditions related to chilling for each site. Here we show species-level estimates from our model (Fig. **??**) for the two most common species in the OSPREE database: *Betula pendula* (A, B) and *Fagus sylvatica* (C, D), for sites that highlight the diversity of possible budburst responses to warming (Fig. S6, which shows general trends across many sites in Central Europe). In some sites, warming increases total chilling estimates (A, C) leading to greater advances in budburst (compared to forcing alone), whereas warming decreases total chilling estimates in other sites (B, D), leading to smaller advances and, eventually, delays with substantial warming. See Supplemental Materials, especially Figs. S6 - S10, for details. Compare this to Fig. 4 in the main text.

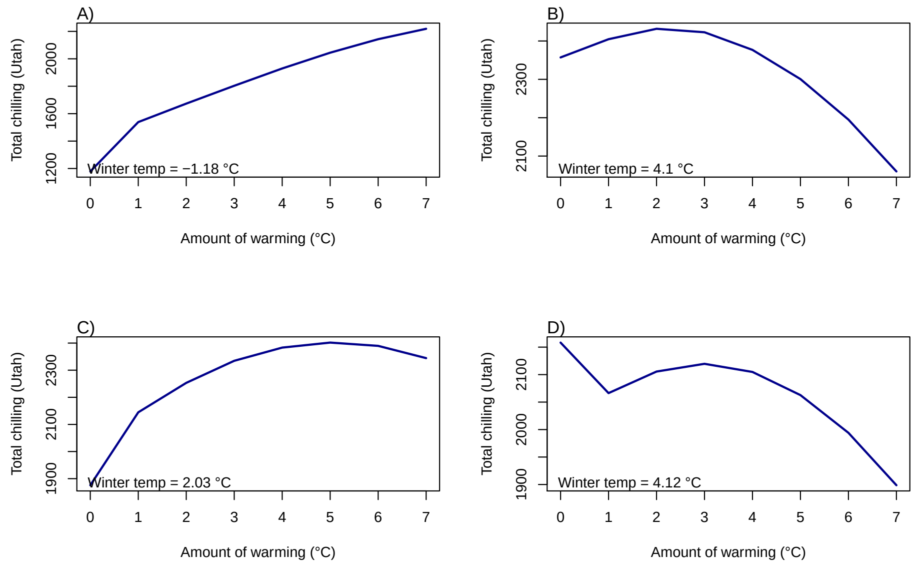

Figure S10: **Implications of global warming on chilling vary by site**, depending on pre-warming climate. At sites in A (46.8°N, 12.8°E) and C (48.8°N, 15.4°E), chilling increases with warming, whereas chilling decreases with warming for the sites in B (48.3°N, 15.8°E) and D (46.7, 15.8°E). Compare to Fig. **??** and Fig. 3 in the main text.

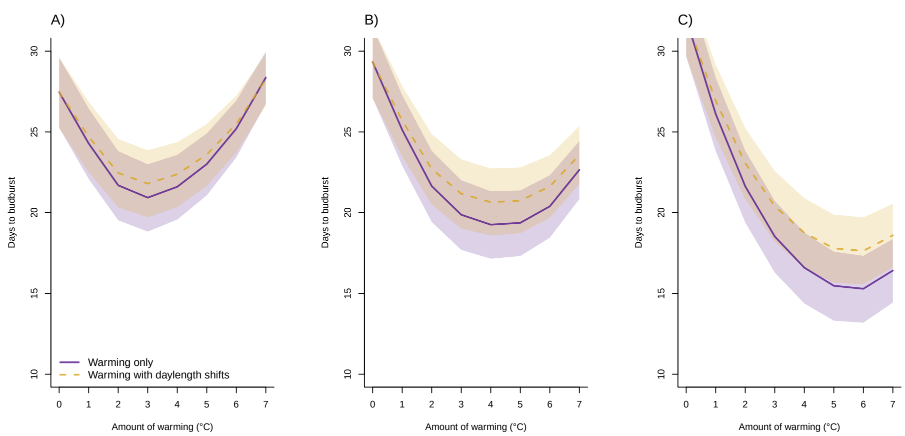

Figure S11: **Budburst is affected by climate-change induced shifts in photoperiod, especially at high latitudes**, though effects vary by site and are minor compared to effects of warming. We show forecasted effects of varying levels of warming on *Fagus sylvatica*, the most photoperiod-sensitive species in our database, across three latitudes within its range, as predicted by the latitude model. The low latitude site (A) is located at 46.8°N, 15.7°E; the mid-latitude site (B) is located at 47.7°N, 16.3°E; and the high-latitude (C) site is located at 48.8°N, 15.4°E.

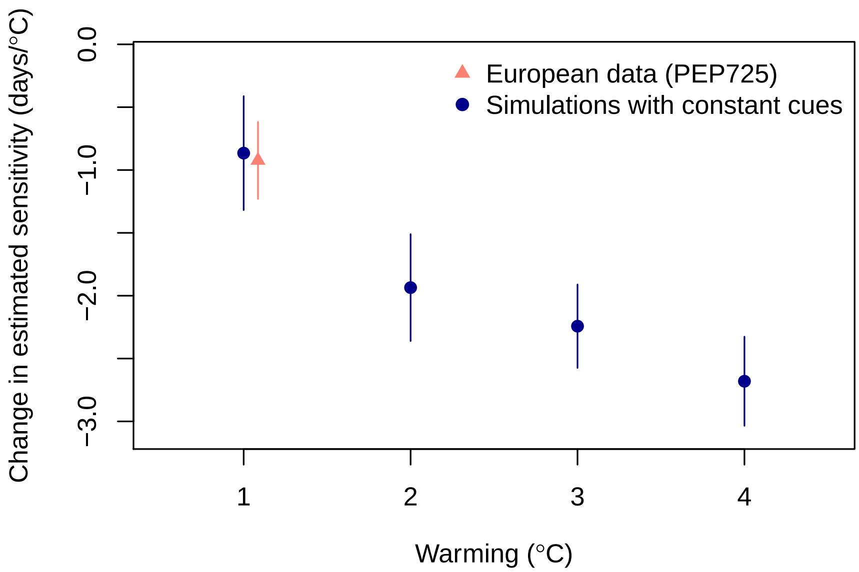

Figure S12: **Declining sensitivities observed in long-term European data for a suite of common trees may be explained by a statistical artifact.** We compared the sensitivity estimated from linear regressions of day of leafout versus mean spring temperature (estimated thus as days/°C) from PEP725 data 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°C ('Simulations') and found the estimated temperature sensitivity measured as days/°C declined even though the underlying cues had not changed, see *Understanding declines in temperature sensitivity in European long-term data* for further details.

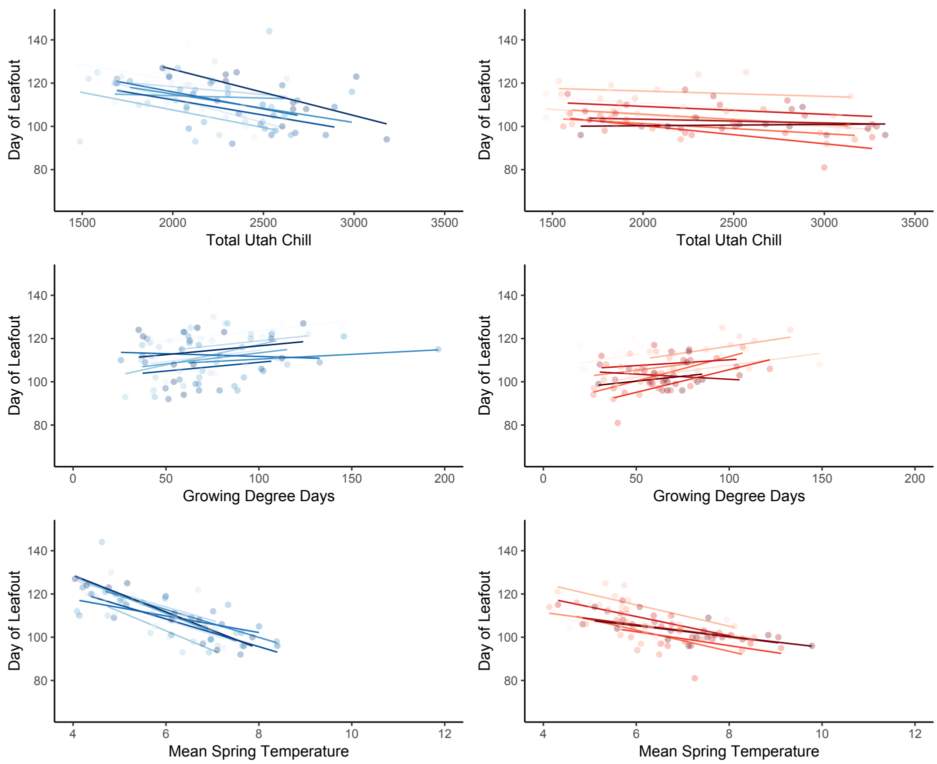

Figure S13: **Day of leaf out versus chilling, growing degree-days, and mean spring temperature** pre- (left panels, 1951-1960) and post-warming (right panels, 2000-2010) for PEP725 sites in Germany where *Betula pendula* phenology has been monitored for decades.

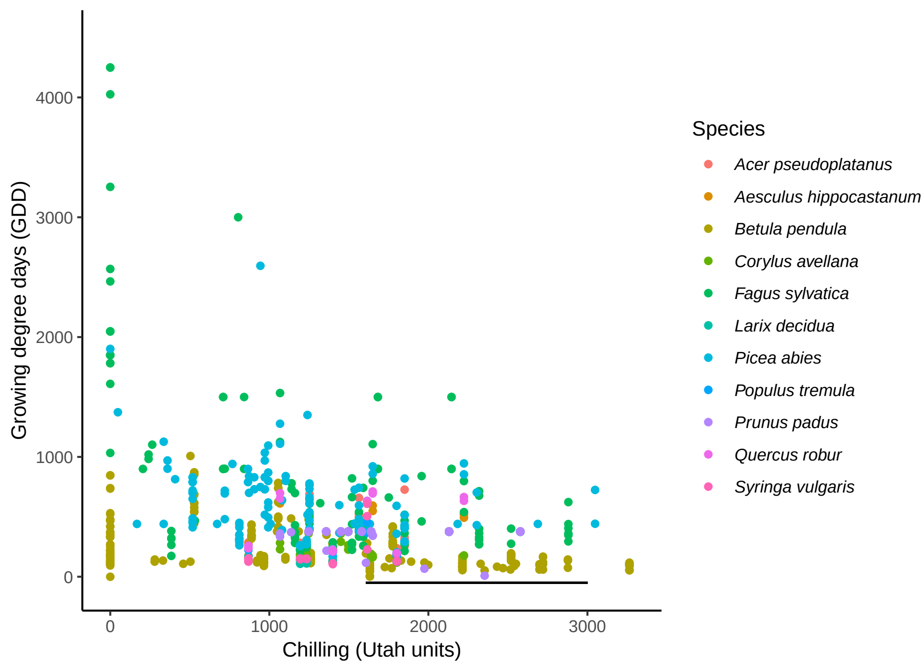

Figure S14: **Growing degree days (GDD) versus chill units at the time of budburst** from the OSPREE database for common species in the PEP725 long-term phenological database. The black line shows the range of chilling (10-90% quantiles) accumulated from 1 September to 1 March for 45 sites for *Betula 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.