Check for updates

Global Ecology and Biogeography

WILEY

Global Ecology and Biogeography A Journal of Macroecology

PERSPECTIVE

Thermal Forcing Versus Chilling? Misspecification of Temperature Controls in Spring Phenology Models

Xiaojie Gao[1,2] | Andrew D. Richardson[3,4] | Mark A. Friedl[5] | Minkyu Moon[6] | Josh M. Gray[1,7]

[1]Center for Geospatial Analytics, North Carolina State University, Raleigh, North Carolina, USA | [2]Harvard Forest, Harvard University, Petersham, Massachusetts, USA | [3]Center for Ecosystem Science and Society, Northern Arizona University, Flagstaff, Arizona, USA | [4]School of Informatics, Computing, and Cyber Systems, Northern Arizona University, Flagstaff, Arizona, USA | [5]Department of Earth and Environment, Boston University, USA | [6]Department of Environmental Science, Kangwon National University, Chuncheon, Korea | [7]Forestry and Environmental Resources, North Carolina State University, Raleigh, North Carolina, USA

Correspondence: Xiaojie Gao (xiaojiegao@fas.harvard.edu)

Received: 5 March 2024 | **Revised:** 7 October 2024 | **Accepted:** 13 October 2024

Handling Editor: Anne Bjorkman

Funding: This work is mainly funded by the Future Investigators in NASA Earth and Space Science and Technology (FINESST) programme (#80NSSC21K1611) and the NASA Grant (80NSSC21K1974). ADR also acknowledges support from EF-1702697, DEB-1637685 and DEB-1832210. We would like to thank the MuSLI team at Lund University for giving us constructive suggestions throughout the analysis. We also want to thank the developers of the R programming language and the 'phenor' package. Finally, we would like to thank the reviewers for the constructive comments, which substantially improved the quality of the manuscript.

Keywords: chilling | phenology model | plant phenology | process-based model | spring leaf emergence | thermal forcing

ABSTRACT

Background: Climate-change-induced shifts in the timing of leaf emergence during spring have been widely documented and have important ecological consequences. However, mechanistic knowledge regarding what controls the timing of spring leaf emergence is incomplete. Field-based studies under natural conditions suggest that climate-warming-induced decreases in cold temperature accumulation (chilling) have expanded the dormancy duration or reduced the sensitivity of plants to warming temperatures (thermal forcing) during spring, thereby slowing the rate at which the timing of leaf emergence is shifting earlier in response to ongoing climate change. However, recent studies have argued that the apparent reductions in temperature sensitivity may arise from artefacts in the way that temperature sensitivity is calculated, while other studies based on statistical and mechanistic models specifically designed to quantify the role of chilling have shown conflicting results.

Methods: We analysed four commonly used combinations of phenology and temperature datasets obtained from remote sensing and ground observations to elucidate whether current model-based approaches robustly quantify how chilling, in concert with thermal forcing, controls the timing of leaf emergence during spring under current climate conditions.

Results: We show that widely used modeling approaches that are calibrated using field-based observations misspecify the role of chilling under current climate conditions as a result of statistical artefacts inherent to the way that chilling is parameterised. Our results highlight the limitations of existing modelling approaches and observational data in quantifying how chilling affects the timing of spring leaf emergence and suggest that decreasing chilling arising from climate warming may not constrain near-future shifts towards earlier leaf emergence in extra-tropical ecosystems worldwide.

1 | Introduction

The timing of leaf emergence, leaf development, and flowering (spring phenology) in vegetation plays an important role in ecosystem function and biosphere–atmosphere exchanges of carbon, energy and water (Korner and Basler 2010; Piao et al. 2019; Richardson et al. 2013). Because spring phenology is widely understood to be controlled by a limited set of bioclimatic factors, shifts in the timing of phenological events provide a sensitive indicator of climate change-induced modifications to ecosystem functions (Gao et al. 2021; Keenan et al. 2014; Korner and Basler 2010; Piao et al. 2019). A large body of both model-based and observational evidence indicates that spring phenology has significantly advanced in many extra-tropical ecosystems in recent decades due to global warming (Keenan et al. 2014; Piao et al. 2019). Such phenological shifts have important consequences for large-scale climate–ecosystem interactions.

Temperature is the dominant control of spring phenology in extra-tropical ecosystems that are not water limited (Chuine and Régnière 2017; Korner and Basler 2010; Richardson et al. 2018). The assumption that spring phenological events are triggered after accumulating sufficient thermal forcing units has been accepted for three centuries (Chuine and Régnière 2017; Réaumur 1735). This assumption is consistent with widespread shifts towards earlier spring phenology under climate warming. At the same time, both empirical and model-based studies have suggested that plants need sufficient cold temperatures (chilling) to break dormancy and that the amount of thermal forcing required to initiate spring phenology is conditioned by the amount of chilling accumulated (Baumgarten et al. 2021; Chmura 2006; Fu, Piao, et al. 2015; Heide 1993; Lin et al. 2022). In recent years, a growing body of observational studies has suggested that the sensitivity of spring phenology to warming temperatures is decreasing, ostensibly as a result of warming-induced decreases in chilling accumulation (Fu, Zhao, et al. 2015; Meng et al. 2020), while other studies have argued that apparent declines in temperature sensitivity may simply be an artefact generated by non-stationarity (Ettinger et al. 2020; Keenan, Richardson, and Hufkens 2020) or nonlinearity (Wolkovich et al. 2021) in the response of spring phenology to thermal forcing. Controlled lab experiments in greenhouse chambers supported the role of chilling (Baumgarten et al. 2021; Ettinger et al. 2020; Lin et al. 2022). However, questions exist regarding whether experimental results accurately reproduce phenological responses to natural variability (Wolkovich et al. 2012), and results from studies using data collected in the field are inconsistent, with some studies performed using the same dataset yielding divergent conclusions (Basler 2016; Fu, Zhao, et al. 2015; Hänninen et al. 2019; Richardson et al. 2009; Wang et al. 2022; Zhang et al. 2022).

This lack of consistency across studies suggests that understanding of what controls the timing of spring phenology is incomplete. Specifically, we do not know when plants start responding to temperatures, what range of temperatures influences spring phenology, and how and why these controls vary by species or climate conditions. Observational studies are limited by the fact that chilling and thermal forcing are not independent, which can lead to spurious results. Similarly, studies that use process-based models rely on robust model formulations and require sufficient data to calibrate model parameters. Indeed, incorrect models can generate satisfactory fits to data (Hunter and Lechowicz 1992), which limits the power of model-based intercomparisons for drawing inferences about phenological processes. Hence, the extent to which observational studies and process-based models can be used to understand the role of chilling in spring phenology is unclear. Consequently, fundamental questions regarding whether or not decreased chilling arising from warming conditions affected the rate of advancement in spring phenology (or even reversed the direction of change) are under debate (Ettinger et al. 2020; Fu, Zhao, et al. 2015; Wang et al. 2020, 2022; Zhang et al. 2018, 2022). To investigate this problem, here we examine commonly used methodologies with four combinations of phenology and temperature datasets to comprehensively elucidate whether current widely used modelling approaches robustly quantify how chilling, in concert with thermal forcing, controls the timing of spring leaf emergence under current climate conditions.

2 | Results and Discussion

2.1 | The Negative Chilling—forcing Relationship

Lack of understanding regarding how to best measure chilling accumulation makes it hard to directly investigate the chilling effect on spring phenology. For this reason, observational studies often calculate chilling accumulation based on a commonly used start date (e.g., 1 November) and temperature threshold (e.g., 0°C or 5°C), and use a linearised negative chilling–forcing relationship to estimate chilling importance (Fu, Piao, et al. 2015; Wang et al. 2020, 2022; Zhang et al. 2018). However, this approach does not account for the fact that chilling and thermal forcing are often correlated (e.g., in most extra-tropical climate zones, colder temperatures, which increase chilling, lead to less thermal forcing and vice versa). Here, we demonstrate that this negative chilling–forcing relationship occurs regardless of whether or not chilling influences spring phenology, which generates an apparent, and potentially spurious, role for chilling under many natural conditions.

To illustrate this effect, we used the Pan European Phenological database (Templ et al. 2018) (PEP725; Table A1, Figure S1c) and the gridded E-OBS temperature dataset (Cornes et al. 2018) to estimate the chilling–forcing relationship using phenology data simulated from a process-based model that excludes chilling. The PEP725 data were collected by citizen scientists at thousands of locations in central Europe over the past several decades and have been widely used to study how climate change affects phenology (Basler 2016; Fu, Zhao, et al. 2015; Wang et al. 2022; Zhang et al. 2022). We used 221,563 individual observations collected at 1177 sites from the PEP725 database (Table A1), representing six dominant plant species, and preprocessed the phenology and temperature datasets following methods used in previous studies (Basler 2016; Fu, Zhao, et al. 2015; Wang et al. 2022; Zhang et al. 2022). For the simulation experiment, we applied the widely used thermal time (TT) model (Réaumur 1735) (see Supporting Information S1) to simulate start of season (SOS) dates that are comparable to the timing observed in the PEP725 dataset. Note that the TT model assumes that spring phenology occurs when a fixed threshold in thermal forcing units (growing degree days, GDDs) has accumulated and does not include chilling. We prescribed a start date (t_0) of 30 January and a base temperature (T_{base}) of 5°C, both of which are commonly used in previous studies, and then estimated a set of SOS dates for each PEP725 site by fitting the TT model. The resulting dataset provides a set of simulated SOS dates that do not include a chilling effect. We then used the simulated SOS dates and temperature data to estimate a chilling–forcing relationship for each site following the methodology used in multiple recent studies (Fu, Zhao, et al. 2015; Wang et al. 2020, 2022) (Figure 1).

Results from this analysis show that even though no chilling was used to simulate SOS dates, 65% of the 1177 sites showed significant negative chilling–forcing relationships in the simulated data (Figure 1a), about the same proportion when using the original SOS dates in the PEP725 dataset (65.6%, Figure S4). Although the magnitude of significance may vary depending on the datasets analysed and the calculations, results from these simulations indicate that the importance of chilling cannot be inferred by fitting this relationship. Stated another way, because the simulated SOS dates were generated solely as a function of thermal forcing, the apparent effect of chilling is spurious. Further, the simulated dataset reproduced significant temporal trends noted in previous studies (Fu, Piao, et al. 2015; Zhang et al. 2022), suggesting continued advancement in the timing of spring phenology from warming (Figure S3a,b) despite declining chilling accumulations (Figure S3c,d) during the past several decades. These results suggest that the apparent impact of changes in chilling on recent spring phenological shifts that have been reported in previous studies was probably misattributed, that is, the methodology tends to identify a significant chilling effect even when chilling does not play a significant role.

To analyse the universality of the negative chilling–forcing relationship we describe above, we conducted a similar simulation experiment using in situ temperature data from the Harvard Forest long-term ecological research site in Central Massachusetts. In this experiment, we removed phenological variation by prescribing all interannual SOS dates to have a single randomly selected value. We then used randomly selected combinations of base temperature and chilling and forcing accumulation start dates to calculate a chilling–forcing relationship using the Harvard Forest temperature data. This experiment was repeated 1000 times (i.e., 1000 random combinations of SOS, base temperature, and chilling and forcing accumulation start dates) and is designed to test whether the chilling–forcing relationship in a different 'chilling-free' SOS dataset is sensitive to the used model parameters. Results from these simulations show that the negative chilling–forcing relationship was significant in 63.8% of the 1000 iterations (Figure S5). This result suggests that the correlation between chilling and forcing is not a by-product of a specific parameter set but is a universal consequence of natural variation in air temperature. Surprisingly, we obtained statistically significant inverse exponential relationships with some parameter combinations (e.g., Figure 1b) that resemble results from the study that originally proposed the chilling–forcing relationship from observations (Cannell and Smith 1983).

The results from this simulation experiment once again demonstrate that a negative chilling–forcing relationship can be obtained, even when chilling exerts no control on SOS and reinforces the conclusion that previous studies may have misidentified the importance of chilling. To provide an intuitive illustration, if we use commonly prescribed values for start date (DOY = 1) and the mean daily temperature used to prescribe base temperature for GDD ($\geq$ 5°C) and chilling days (< 5°C) (e.g., Cannell and Smith 1983; Chen et al. 2018; Fu, Zhao, et al. 2015) and assume that SOS happens on DOY 130 (representative of Harvard Forest), then, by definition, chilling and forcing are negatively correlated. Complementary to

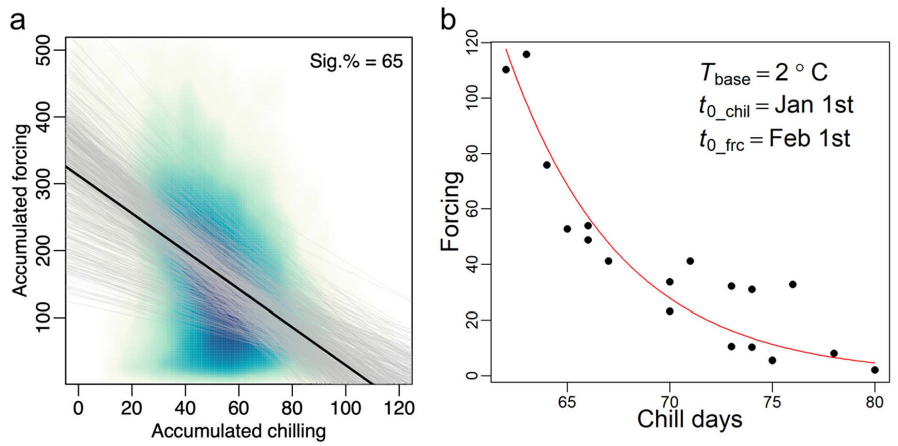

FIGURE 1 | Negative chilling–forcing relationship calculated from simulated start of season (SOS) dates by the thermal time (TT) model. (a) Accumulated thermal forcing versus accumulated chilling estimated from observed temperature data. The accumulated chilling was calculated as the number of chill days when the daily mean temperature was between 0°C and 5°C from 1 November of the previous year to the average SOS date, and forcing was calculated as growing degree days with T_{base} = 5°C from t_0 = 30 to the average SOS date. The figure shows the fitted negative chilling–forcing relationship obtained from the simulated data. The grey lines represent statistically significant ($p < 0.05$) site-specific slopes, and the black line is the average slope. (b) Inverse exponential chilling–forcing relationship between chill days and forcing units based on the TT model with spring phenology fixed at DOY 80 for all years. Temperature data were measured at the Harvard Forest Environmental Monitoring Site eddy-covariance tower. Chill days are computed as the number of days with daily mean temperatures below T_{base} from 1 January (t_{0_chil}). Forcing represents the accumulated temperature units above T_{base} from 1 February (t_{0_frc}).

the findings identifying non-stationarity (Ettinger et al. 2020; Keenan, Richardson, and Hufkens 2020) and nonlinearity (Wolkovich et al. 2021) as sources of error in phenological models, our results indicate that regardless of how the phenological sensitivity to thermal forcing has changed, the inherent correlation between chilling and thermal forcing may prevent accurate attribution of the apparent changes in SOS to chilling dynamics. In addition, chilling has been reported to be increased or decreased by climate warming (Wang et al. 2020; Chen, Wang, and Inouye 2017). Thus, the magnitude of chilling–forcing correlation may vary depending on the local climate conditions and the specific method used to calculate chilling. We argue that the lack of consensus on how to best quantify chilling is a key obstacle that is preventing the community from disentangling natural correlations from mechanisms, especially since the requirement, magnitude and responsive temperatures of chilling may vary among species, genotypes and/or geography (Baumgarten et al. 2021; Chmura 2006; Heide 1993; Lin et al. 2022; Rousi and Pusenius 2005; van der Schoot and Rinne 2011; Vitasse et al. 2009). Note that our analysis does not disqualify the role of chilling found in experimental studies, since chilling can be an important process but may be easily fulfilled under current climate conditions in many regions and thus has little effect on SOS (Basler 2016; Ettinger et al. 2020; Korner and Basler 2010).

2.2 | Deriving Process From Process-Based Models

Process-based spring phenology models are widely used to study mechanic variation across species in their response and sensitivity to thermal and other types of forcing (Basler 2016; Chuine 2000; Hufkens et al. 2018; Hunter and Lechowicz 1992), although the widely used inverse modelling approach solely based on SOS observations for model parameter calibration is increasingly questioned (Chuine et al. 2016; Chuine and Régnière 2017; Hänninen et al. 2019). To investigate sources of disagreement regarding the importance of chilling among model-based observational studies, here we use four combinations of commonly used phenology and temperature datasets (Table A2) to test five widely used process-based models that use thermal forcing and chilling to predict the timing of spring phenology: the TT model (Réaumur 1735), the Parallel (PA) model (Hänninen 1990), the Sequential (SQ) model (Hänninen 1990; Kramer 1994), the Alternating (AT) model (Kramer 1994) and the Unified (UN) model (Chuine 2000). The TT model does not include chilling, the PA, SQ, AT and UN models encode chilling in different ways, and the UN model is a generalised model for which all of the other models are special cases (Table A2; Figure S1, Table S1). To provide a baseline comparison, we also considered a simple linear regression (LIN) model in which the timing of spring phenology is modelled as a linear function of preseason mean temperature (Basler 2016). We tested single monthly mean temperatures from January to May as well as mean temperatures across months from January–March and March–May for the preseason duration in the LIN model and retained the best-fitting variant for comparison with the process-based models. In total, six models were applied to the four combinations of phenology and temperature data datasets. For details on the various models, please see the Methods in Appendix A and the Supporting Information S1.

We used the models and datasets described above to explore two questions: (1) Do the model results support the importance of chilling? and (2) Do model parameters quantify information that can be used to infer phenological responses to future climate change? Depending on the nature of each data source, we conducted pooled analysis, and site- and species-specific analyses. The pooled analyses fit individual models using all available data across site years in each dataset (i.e., one model for each dataset, except the PEP725 phenology dataset due to its size), while the site- and species-specific analyses fit models separately for each site and/or species (i.e., one model for each site or species in each dataset). Both goodness-of-fit and cross-validation were used to evaluate model performance. Leave-one-site-out cross-validation was applied to the pooled analyses, while leave-1-year-out cross-validation was applied to the site-specific analyses. In the leave-one-site-out cross-validation, the fitted SOS dates for individual sites were either pooled to evaluate model performance in capturing spatial variation across sites or converted into site-specific anomalies to evaluate model performance in capturing interannual variation. Models were evaluated using R^2 and root mean square error (RMSE) values calculated from linearly regressing observed (Obs.) SOS dates against estimated (Est.) SOS dates. Analysis of variance of RMSE values was conducted to test the significance of differences between model fits for site-specific analyses.

Our results show that although goodness-of-fit metrics exhibited significant variability for the same phenology and temperature data (Figure S6), all models performed similarly in cross-validation (Figure 2; Figure S7). In the pooled analyses, the performance of models did not significantly differ in capturing spatial (Figure S7) and interannual phenology variations (Figure 2). Interestingly, the LIN model achieved nearly comparable performance to the process-based models (Figure 2; Figure S6), which has also been noted in previous studies (Basler 2016; Hänninen et al. 2019; Olsson and Jönsson 2014).

To evaluate uncertainty in parameter estimates for the process-based models, we employed bootstrapping using 1000 iterations to resample with replacement from all available data, that is, the resampled dataset had the same number of site years as the original dataset. Then, for each iteration, we estimate model parameters by fitting the models with the resampled dataset. We used the results from this analysis to estimate 95% confidence intervals (CIs) for each estimated parameter.

Results from bootstrapping show that the estimated parameters in the process-based models have large uncertainty (Figure 3; Figures S11—S16). For example, the 95% CI for base temperature (T_{base}) was −3°C to 5°C for the TT model, the 95% CI for the estimated start day of chilling accumulation ($t_{0-\mathrm{chil}}$) for the PA model ranged from mid-October to 1 March, and the 95% CI for minimum growth competence (C_{min}) in the PA model varied from 0 to 0.6, which covers more than half of its valid range ($C_{\mathrm{min}} \in [0, 1]$). In short, these results show that highly dissimilar parameter values can result in model predictions with very similar performance relative to observations.

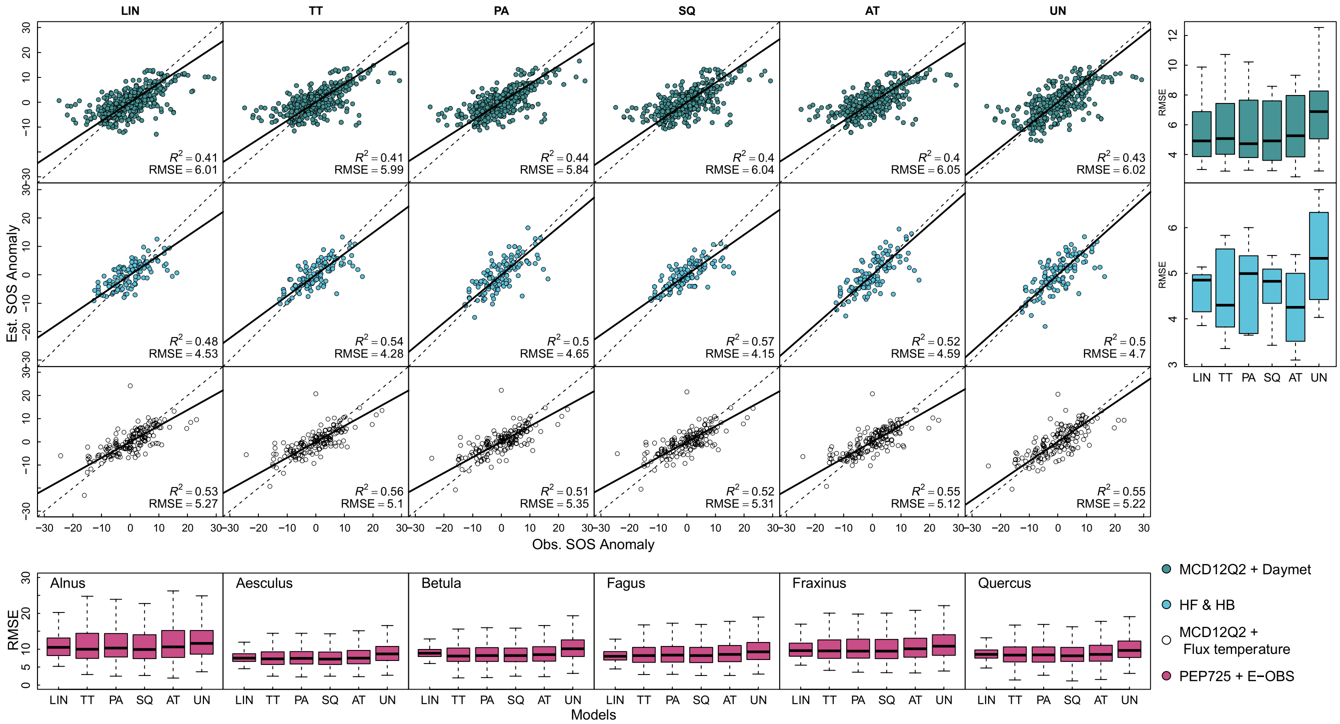

FIGURE 2 | Cross-validation results for spring phenology models: linear regression (LIN), Thermal Time (TT), Parallel (PA), Sequential (SQ), Alternating (AT), and Unified (UN). Scatter plots show results from pooled analysis and boxplots show results from site- or species-specific analyses. Black lines in the scatter plots are Type-II regression lines accounting for errors in both the response and predictor variables. The first row shows results using the MCD12Q2 phenology product derived from the Moderate Resolution Imaging Spectroradiometer and the Daymet temperature dataset at flux tower locations. The second row shows results using ground phenological observations obtained at Harvard Forest (HF) and Hubbard Brook Experimental Forest (HB) along with temperatures measured at each site. The third row shows results using the MCD12Q2 phenology product and ground temperatures measured at flux tower sites. Note that site-specific model comparisons were excluded from the analysis based on MCD12Q2 data due to insufficient site years at each site. The fourth row shows species-specific results using the PEP725 phenological observations and E-OBS temperatures.

While our result that the different models perform similarly in cross-validation may be surprising, it is consistent with previous studies (Basler 2016; Hufkens et al. 2018; Hunter and Lechowicz 1992; Melaas et al. 2016; Olsson and Jönsson 2014; Richardson and O'Keefe 2009). However, our results are based on a larger and more diverse set of data sources and a more consistent analytical design. More importantly, these results identify three fundamental challenges in assigning process-based interpretations to spring phenology model results.

First, when fitting to observational data, the more complex models effectively reduce to the TT model even though they parameterise processes that trigger spring phenology differently. More specifically, we found that the estimated chilling parameters in the PA, AT and UN models had negligible impact on forcing accumulation. As a consequence, the forcing trajectory and required forcing units were effectively identical to the TT model. Similarly, when the chilling fulfilment date is close to the start date of thermal forcing accumulation, the SQ model simplifies to the TT model. In theory, the UN model should be the most accurate because it is designed to flexibly account for both chilling and forcing processes. While it did achieve the best performance in many goodness-of-fit results (Figure S6), its performance was similar to other models in cross-validation (Basler 2016; Chuine 2000; Hufkens et al. 2018). Again, it is possible that chilling requirements are easily fulfilled under natural conditions in many locations, and so limited variability in chilling makes the chilling process difficult to identify among the different models. However, testing this hypothesis is difficult because it would require data obtained under rare conditions such as extremely warm winters or by manipulative experiments, which are not widely available.

Second, sample size and differences in the number of parameters across models need to be carefully considered in model intercomparisons. Specifically, because the response of phenological processes to thermal forcing and chilling likely varies by species and location, fitting phenological models to species- or site-specific data is generally preferred. Also, since the practice of process-based model calibration is computationally intensive, goodness-of-fit and the Akaike information criterion (AIC/AICc) are commonly used in place of cross-validation when intercomparing models. Given the relatively small number of site- or species-specific phenological observations compared to the number of parameters in process-based models, AICc is generally preferred (Equation A1). Using the TT model ($k = 3$) as the benchmark, we found that the statistical support for a more complex model ($\Delta AICc < -2$) requires the model to have a substantially lower RMSE or many more years of observations than are typical of current phenological datasets (Figure 4). For example, with 20 years of phenology

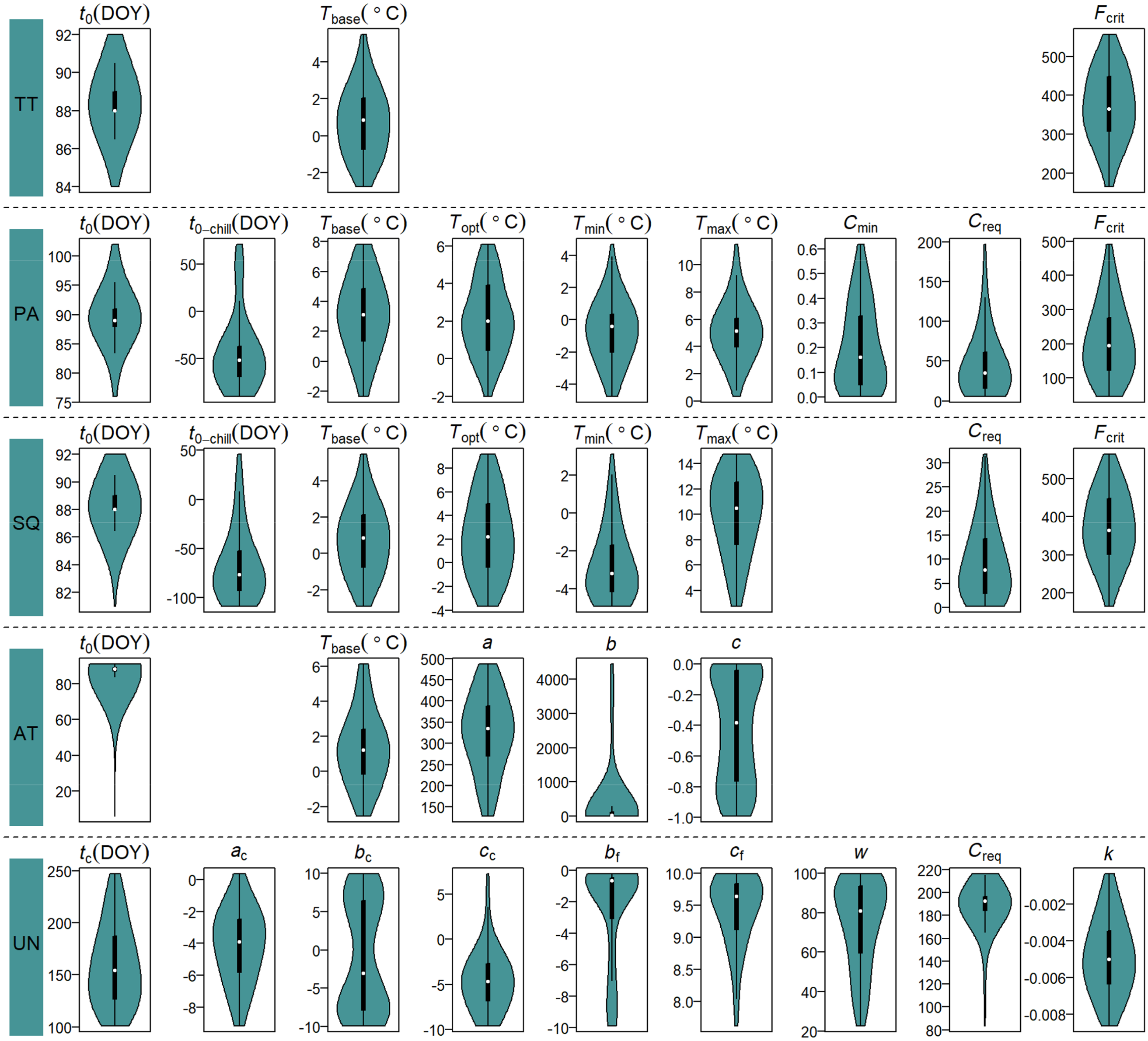

FIGURE 3 | 95% confidence intervals for model parameters from the bootstrapping analysis. Violin plots show the density distributions for estimated parameters. The inner box plots show median values, first and third quantiles, and 1.5 times the inter-quantile range. Values outside of the 95% confidence intervals were removed from the graphs. For each case, the maximum number of iterations for the optimisation algorithm to converge was set to 100,000.

observed at a site, the 5-parameter AT model would need to reduce the RMSE by at least 20% to outperform the 3-parameter TT model. Similarly, to support the 9-parameter PA model, the minimum required reduction in RMSE is 54%, which is more than half of the TT model's RMSE. Equally important, given the revisit frequency of many phenological observations, the inherent uncertainty in the data is substantial, which makes further RMSE reduction challenging. In this context, the ΔAICc result (Figure 4) both confirms our cross-validation results and likely explains the reason why previous studies failed to differentiate among models.

Third, parameter correlation is a major source of uncertainty that complicates interpretation of model-based results. For example, the correlation between T_{base} and F_{crit} in the TT, PA, and SQ models is close to 1.00 (Figure 5), and the correlation between w and k in the UN model is −0.64 (Figure S15). Because of this, it is effectively impossible to estimate the 'true' parameter values, regardless of the number or quality of observations or the statistical optimisation algorithm. A commonly used practice is to prescribe T_{base} to be 0°C or 5°C (Cannell and Smith 1983; Kaduk and Los 2011). However, seasonality and variability in temperatures introduce large uncertainty in the appropriate start date (t_0) regardless of the date on which accumulation begins (Linkosalo, Lappalainen, and Hari 2008). Some studies have prescribed values for both t_0 and T_{base} for relatively small and homogeneous areas (Cannell and Smith 1983; Kaduk and Los 2011; Melaas et al. 2016), but, as we show above, this is effectively equivalent to LIN, and more importantly, it is likely that these parameters vary as a function of location and/or species (Fang et al. 2022; Fu et al. 2012; Hänninen et al. 2019). In short, our results suggest that current process-based spring phenology models and parameters calibrated by observational data may not support more complex models that include parameterisations for chilling processes.

3 | Implications and Suggestions for Future Studies

Our results indicate that the importance of chilling in controlling spring phenology in plants under current climate conditions has probably been misspecified in many observational studies. This conclusion is consistent with lab experiments indicating that even though chilling is an important cue, thermal forcing dominates the observed phenological trends under current environmental conditions (Ettinger et al. 2020). Thus, accounting for chilling may not significantly improve the performance of these models. Interestingly, as our analysis shows, observational studies that estimate the chilling–forcing relationship and intercompare process-based models can generate contradictory results regarding whether chilling significantly affects spring phenology, even when using the same dataset (e.g., the PEP725 dataset). Specifically, we show that the negative chilling–forcing relationship commonly used to support the role of chilling can be reproduced, with a similar significance level, by a process that excludes chilling under natural temperature conditions (Figure 1). Critically, this suggests that results from recent studies based on empirical chilling–forcing relationships estimated from natural phenological observations may be statistically spurious, and by extension, do not provide evidence that reduced chilling has lowered the sensitivity of spring phenology to temperature. Similarly, results from our intercomparison of process-based models do not support the role of chilling in explaining variability in spring phenology observations. Given the relatively large uncertainty in model parameters and commonly used goodness-of-fit metrics, model intercomparison can generate results that appear to identify differences among models that are actually caused by small differences in data that are unrelated to phenological processes. That said, it is important to note that the results we present are not intended to discount the validity of widely used process models, nor do they necessarily imply errors in the way they parameterise processes. Rather, our results underscore the limitations of observational data that are commonly used in these studies, and by extension, the limitations imposed on conclusions that can be drawn from them based on commonly used phenological models. Specifically, currently available observational data may not provide a sufficient basis for selecting more complex models over simpler models or for unequivocally identifying or quantifying ecological processes using parsimonious explanations (Figure 4). Hence, our results help explain the diverse and often conflicting results obtained from observational studies attempting to identify mechanisms by fitting statistical and process-based

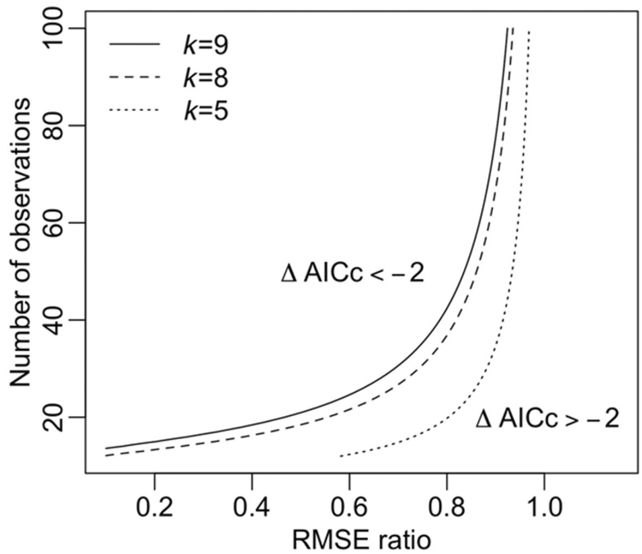

FIGURE 4 | Variation in the small-sample-corrected Akaike information criterion difference (ΔAICc) as a function of root mean square error (RMSE) and the number of observations in the dataset used to compare process-based spring phenology. The curves represent ΔAICc variations when intercomparing more complex models with k parameters to a simple benchmark model ($k = 3$).

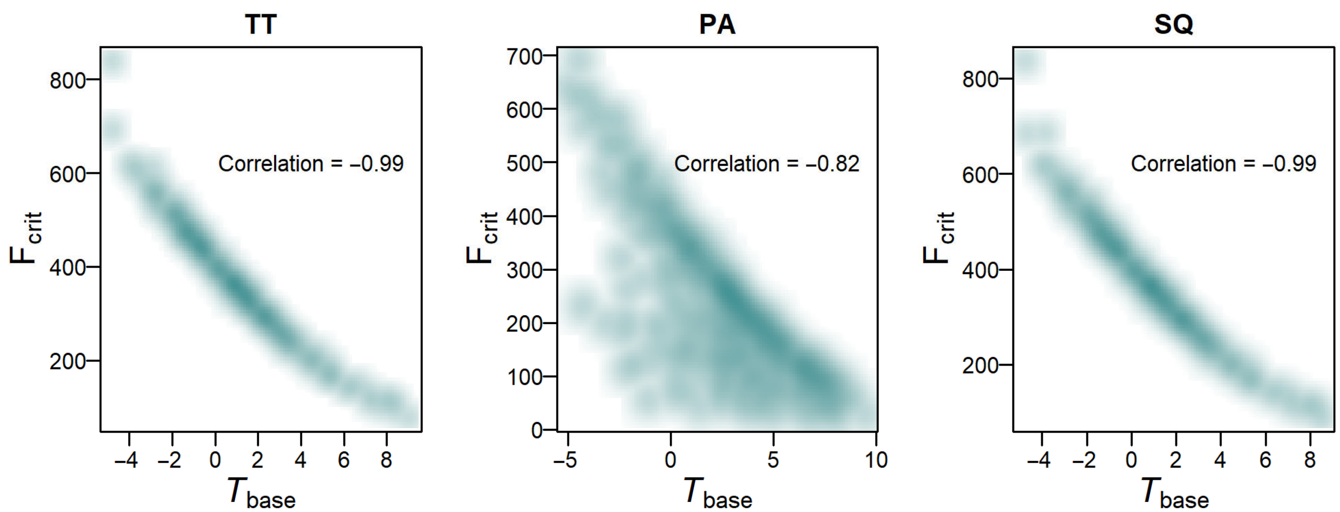

FIGURE 5 | Correlation between the forcing threshold parameter (F_{crit}) and base temperature (T_{base}) in the Thermal Time (TT) model, the Parallel (PA) model and the Sequential (SQ) model. Darker colours represent higher density of data points. Data were obtained from the bootstrapping analysis.

models, quantify temperature sensitivity and predict changes in spring phenology as the climate becomes warmer.

The challenges that we identify in accurately quantifying processes and estimating parameters may further account for the poor performance when model results are extrapolated. For example, models calibrated using one dataset often do not accurately capture variability in spring phenology in independent datasets, even though both datasets include the same species and are from nearby geographical regions (Chuine, Cour, and Rousseau 1998; Chuine and Régnière 2017; Linkosalo, Lappalainen, and Hari 2008). To rigorously test hypotheses regarding the impact of chilling on spring phenology, controlled experiments that disrupt the correlation between winter and spring temperatures are critical (Baumgarten et al. 2021; Ettinger et al. 2020; Lin et al. 2022; Polgar and Primack 2011). However, manipulative experiments are often limited to restricted geographical regions and time periods. Field observations from unusual years or phenomena such as urban heat islands and coastal effects, which elevate winter temperatures relative to surrounding environments, may serve as natural experiments. However, caution is required to account for co-varying factors aside from temperature (e.g., variation in species, location adaptations, etc.). In this context, it is important to note that our analysis did not include models incorporating photoperiod effects. While photoperiod may enhance model performance in pooled analyses, because it has no interannual variability in site-specific observations, excluding it does not affect our intercomparison results or conclusions from temperature-driven models.

Similar to our comments above regarding the role of chilling, we do not want our results to be interpreted as discounting the utility of models in understanding and forecasting phenological dynamics. Indeed, we believe that mechanistic plant phenology models play a crucial role in predicting how plants will respond to potential climate changes in the future. However, to enhance their effectiveness, we propose three key improvements.

First, interpretation of results from these models needs to carefully differentiate mechanisms from correlations. In addition to standard cross-validation and comparisons with independent datasets, rigorous consideration of parameter uncertainty and process simulation are crucial (Hänninen et al. 2019; Hunter and Lechowicz 1992), especially when making ecological inferences. Given the complexity of model structures and natural correlations in ecological data, hidden statistical artefacts can generate apparently novel and exciting results that are spurious. In addition to designing controlled experiments to validate results derived from models, bootstrapping and process simulation should be included to identify and reduce spurious results. Recent developments in ecological forecasting (Lewis et al. 2023; Wheeler et al. 2024), which uses simulation but focuses on prediction, may also be useful to test hypothetical mechanisms under novel conditions. In general, scrutiny and scepticism are warranted if a process can be generated by a simulation even without embedding the process, if a result is sensitive to small changes in dataset, and if a proposed mechanism does not produce reasonable predictions.

Second, improved mechanistic representation of phenological processes is needed. Although model complexity needs to be accounted for, integrating more realistic mechanisms is critical. For example, even though the timing of temperature exposure may be significant, most methods do not effectively differentiate between early and late spring warming (Friedl et al. 2014). Similarly, vascular reactivation and development may regulate the sensitivity of leaf buds responding to environmental factors through the transport of water, nutrients and hormones (Savage and Chuine 2021). Thus, process-based models that incorporate more direct time-dependent dynamic phenological processes or that are adapted from time-to-event survival analysis, which quantifies time-varying effects, should be more considered (Clark et al. 2014; Elmendorf et al. 2019; Moon 2021; Templ, Fleck, and Templ 2017; Wheeler et al. 2024; Supporting Information 2.5). In particular, survival analysis can be used to test variables without limiting them to a specific process (e.g., dormancy release, chilling and forcing accumulation) and so may serve as a complementary tool to process-based models for investigating factors that are unclear in terms of importance or mechanisms (Supporting Information S1).

Third, knowledge from controlled experiments should be integrated into modelling. Instead of relying solely on statistical optimisation, incorporating knowledge from field observations and controlled experimental studies may enhance model robustness. These priors can be embedded into models using Generalised Likelihood Uncertainty Estimation (GLUE; Beven and Binley 1992) or Bayesian methods (Reich and Ghosh 2019) while accounting for uncertainty. For example, the dormancy release date, which is critical to phenological modelling (Chuine et al. 2016), may be incorporated into process-based models as prior information using a Bayesian framework. This approach prevents parameter calibration from relying exclusively on optimising model fits to observed SOS observations and accounts for uncertainty propagation in prior knowledge. Similarly, information such as effective temperature ranges and optimal leaf development rates obtained from controlled lab experiments can also be integrated.

In conclusion, using observations collected under current climate conditions, we show that widely used models can generate spurious results regarding the role of chilling in controlling spring phenology. We recommend that future studies be aware of natural correlations in observational data, explore solutions to overcome the limitations of process-based models, use simulations and cross-validation to assess the robustness of models, and be conservative in drawing inferences. We also recommend that future research be explicitly designed using controlled experiments to investigate the internal timing and processes involved in ending dormancy and triggering spring phenology.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

All data used in this study are publicly available. Specifically, the MCD12Q2 phenology product can be downloaded at https://lpdaac.usgs.gov/products/mcd12q2v006/. The Daymet temperature data product can be downloaded at https://daymet.ornl.gov/. The FLUXNET2015 dataset can be downloaded from https://fluxnet.org/data/fluxnet2015-dataset/. The ground-observed phenology and temperature data at Harvard Forest and Hubbard Brook Experimental Forest can be obtained from https://harvardforest.fas.harvard.edu/data-archive and https://hubbardbrook.org/data-catalog/. The PEP725 phenology data and the corresponding E-OBS temperature data can be downloaded from http://www.pep725.eu/ and https://cds.climate.copernicus.eu/cdsapp#!/dataset/insitu-gridded-observations-europe?tab=form. Code scripts used to reproduce the work can be found at: https://zenodo.org/records/13901038.

References

Bailey, A. 2019. "Hubbard Brook Experimental Forest (US Forest Service): Routine Seasonal Phenology Measurements, 1989-Present."

Basler, D. 2016. "Evaluating Phenological Models for the Prediction of Leaf-Out Dates in Six Temperate Tree Species Across Central Europe." *Agricultural and Forest Meteorology* 217: 10–21. https://doi.org/10.1016/j.agrformet.2015.11.007.

Baumgarten, F., C. M. Zohner, A. Gessler, and Y. Vitasse. 2021. "Chilled to Be Forced: The Best Dose to Wake Up Buds From Winter Dormancy." *New Phytologist* 12: 1366–1377.

Beven, K., and A. Binley. 1992. "The Future of Distributed Models: Model Calibration and Uncertainty Prediction." *Hydrological Processes* 6, no. 3: 279–298. https://doi.org/10.1002/hyp.3360060305.

Cannell, M. G. R., and R. I. Smith. 1983. "Thermal Time, Chill Days and Prediction of Budburst in Picea Sitchensis." *Journal of Applied Ecology* 20: 951–963.

Chen, L., J. Huang, Q. Ma, et al. 2018. "Spring Phenology at Different Altitudes Is Becoming More Uniform Under Global Warming in Europe." *Global Change Biology* 24, no. 9: 3969–3975. https://doi.org/10.1111/gcb.14288.

Chen, X., L. Wang, and D. Inouye. 2017. "Delayed Response of Spring Phenology to Global Warming in Subtropics and Tropics." *Agricultural and Forest Meteorology* 234–235: 222–235. https://doi.org/10.1016/j.agrformet.2017.01.002.

Chmura, D. J. 2006. "Phenology Differs Among Norway Spruce Populations in Relation to Local Variation in Altitude of Maternal Stands in the Beskidy Mountains." *New Forests* 32, no. 1: 21–31. https://doi.org/10.1007/s11056-005-3390-2.

Chuine, I. 2000. "A Unified Model for Budburst of Trees." *Journal of Theoretical Biology* 207, no. 3: 337–347. https://doi.org/10.1006/jtbi.2000.2178.

Chuine, I., M. Bonhomme, J.-M. Legave, et al. 2016. "Can Phenological Models Predict Tree Phenology Accurately in the Future? The Unrevealed Hurdle of Endodormancy Break." *Global Change Biology* 22, no. 10: 3444–3460. https://doi.org/10.1111/gcb.13383.

Chuine, I., P. Cour, and D. D. Rousseau. 1998. "Fitting Models Predicting Dates of Flowering of Temperate-Zone Trees Using Simulated Annealing." *Plant, Cell & Environment* 21, no. 5: 455–466. https://doi.org/10.1046/j.1365-3040.1998.00299.x.

Chuine, I., and J. Régnière. 2017. "Process-Based Models of Phenology for Plants and Animals." *Annual Review of Ecology, Evolution, and Systematics* 48, no. 1: 159–182. https://doi.org/10.1146/annurev-ecolsys-110316-022706.

Clark, J. S., C. Salk, J. Melillo, and J. Mohan. 2014. "Tree Phenology Responses to Winter Chilling, Spring Warming, at North and South Range Limits." *Functional Ecology* 28, no. 6: 1344–1355. https://doi.org/10.1111/1365-2435.12309.

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, no. 17: 9391–9409.

Crimmins, T., E. Denny, E. Posthumus, et al. 2022. "Science and Management Advancements Made Possible by the USA National Phenology Network's Nature's Notebook Platform." *Bioscience* 72, no. 9: 908–920.

Elmendorf, S. C., T. M. Crimmins, K. L. Gerst, and J. F. Weltzin. 2019. "Time to Branch Out? Application of Hierarchical Survival Models in Plant Phenology." *Agricultural and Forest Meteorology* 279: 107694. https://doi.org/10.1016/j.agrformet.2019.107694.

Ettinger, A. K., C. J. Chamberlain, I. Morales-Castilla, et al. 2020. "Winter Temperatures Predominate in Spring Phenological Responses to Warming." *Nature Climate Change* 10: 1137–1142. https://doi.org/10.1038/s41558-020-00917-3.

Fang, J., J. A. Lutz, H. H. Shugart, L. Wang, F. Liu, and X. Yan. 2022. "Continental-Scale Parameterization and Prediction of Leaf Phenology for the North American Forests." *Global Ecology and Biogeography 31* 31: 1603–1615. https://doi.org/10.1111/geb.13533.

Friedl, M., J. Gray, and D. Sulla-Menashe. 2019. "MCD12Q2 MODIS/Terra+Aqua Land Cover Dynamics Yearly L3 Global 500m SIN Grid V006. NASA EOSDIS Land Processes DAAC." Accessed January 23, 2020. https://doi.org/10.5067/MODIS/MCD12Q2.006.

Friedl, M. A., J. M. Gray, E. K. Melaas, et al. 2014. "A Tale of Two Springs: Using Recent Climate Anomalies to Characterize the Sensitivity of Temperate Forest Phenology to Climate Change." *Environmental Research Letters* 9, no. 5: 054006. https://doi.org/10.1088/1748-9326/9/5/054006.

Fu, Y. H., M. Campioli, M. van Oijen, G. Deckmyn, and I. A. Janssens. 2012. "Bayesian Comparison of Six Different Temperature-Based Budburst Models for Four Temperate Tree Species." *Ecological Modelling* 230: 92–100. https://doi.org/10.1016/j.ecolmodel.2012.01.010.

Fu, Y. H., S. Piao, Y. Vitasse, et al. 2015. "Increased Heat Requirement for Leaf Flushing in Temperate Woody Species Over 1980–2012: Effects of Chilling, Precipitation and Insolation." *Global Change Biology* 21, no. 7: 2687–2697. https://doi.org/10.1111/gcb.12863.

Fu, Y. H., H. Zhao, S. Piao, et al. 2015. "Declining Global Warming Effects on the Phenology of Spring Leaf Unfolding." *Nature* 526, no. 7571: 104–107. https://doi.org/10.1038/nature15402.

Gao, X., J. Gray, C. W. Cohrs, R. Cook, and T. J. Albaugh. 2021. "Longer Greenup Periods Associated With Greater Wood Volume Growth in Managed Pine Stands." *Agricultural and Forest Meteorology* 297: 108237. https://doi.org/10.1016/j.agrformet.2020.108237.

Hänninen, H. 1990. *Modelling Bud Dormancy Release in Trees From Cool and Temperate Regions* (Issue 213). https://www.cabidigitallibrary.org/doi/full/10.5555/19910652040.

Hänninen, H., K. Kramer, K. Tanino, R. Zhang, J. Wu, and Y. H. Fu. 2019. "Experiments Are Necessary in Process-Based Tree Phenology Modelling." *Trends in Plant Science* 24, no. 3: 199–209. https://doi.org/10.1016/j.tplants.2018.11.006.

Heide, O. M. 1993. "Daylength and Thermal Time Responses of Budburst During Dormancy Release in Some Northern Deciduous Trees." *Physiologia Plantarum* 88, no. 4: 531–540. https://doi.org/10.1111/j.1399-3054.1993.tb01368.x.

Hufkens, K., D. Basler, T. Milliman, E. K. Melaas, and A. D. Richardson. 2018. "An Integrated Phenology Modelling Framework in r." *Methods in Ecology and Evolution* 9, no. 5: 1276–1285. https://doi.org/10.1111/2041-210X.12970.

Hunter, A. F., and M. J. Lechowicz. 1992. "Predicting the Timing of Budburst in Temperate Trees." *Journal of Applied Ecology* 29, no. 3: 597–604. https://doi.org/10.2307/2404467.

Kaduk, J. D., and S. O. Los. 2011. "Predicting the Time of Green Up in Temperate and Boreal Biomes." *Climatic Change* 107, no. 3: 277–304.

Keenan, T. F., J. Gray, M. A. Friedl, et al. 2014. "Net Carbon Uptake Has Increased Through Warming-Induced Changes in Temperate Forest

Phenology." *Nature Climate Change* 4, no. 7: 598–604. https://doi.org/10.1038/nclimate2253.

Keenan, T. F., A. D. Richardson, and K. Hufkens. 2020. "On Quantifying the Apparent Temperature Sensitivity of Plant Phenology." *New Phytologist* 225, no. 2: 1033–1040. https://doi.org/10.1111/nph.16114.

Korner, C., and D. Basler. 2010. "Phenology Under Global Warming." *Science* 327, no. 5972: 1461–1462. https://doi.org/10.1126/science.1186473.

Kramer, K. 1994. "A Modelling Analysis of the Effects of Climatic Warming on the Probability of Spring Frost Damage to Tree Species in the Netherlands and Germany." *Plant, Cell & Environment* 17, no. 4: 367–377. https://doi.org/10.1111/j.1365-3040.1994.tb00305.x.

Lewis, A. S. L., C. R. Rollinson, A. J. Allyn, et al. 2023. "The Power of Forecasts to Advance Ecological Theory." *Methods in Ecology and Evolution* 14, no. 3: 746–756. https://doi.org/10.1111/2041-210X.13955.

Lin, S., H. Wang, Q. Ge, and Z. Hu. 2022. "Effects of Chilling on Heat Requirement of Spring Phenology Vary Between Years." *Agricultural and Forest Meteorology* 312: 108718. https://doi.org/10.1016/j.agrformet.2021.108718.

Linkosalo, T., H. K. Lappalainen, and P. Hari. 2008. "A Comparison of Phenological Models of Leaf Bud Burst and Flowering of Boreal Trees Using Independent Observations." *Tree Physiology* 28, no. 12: 1873–1882. https://doi.org/10.1093/treephys/28.12.1873.

Melaas, E. K., D. Sulla-Menashe, J. M. Gray, et al. 2016. "Multisite Analysis of Land Surface Phenology in North American Temperate and Boreal Deciduous Forests From Landsat." *Remote Sensing of Environment* 186: 452–464. https://doi.org/10.1016/j.rse.2016.09.014.

Meng, L., J. Mao, Y. Zhou, et al. 2020. "Urban warming advances spring phenology but reduces the response of phenology to temperature in the conterminous United States." *Proceedings of the National Academy of Sciences of the United States of America* 117, no. 8: 4228–4233. https://doi.org/10.1073/pnas.1911117117.

Moon, M. 2021. "Using Time Series of MODIS Land Surface Phenology to Model Temperature and Photoperiod Controls on Spring Greenup in North American Deciduous Forests." *Remote Sensing of Environment* 13: 112466. https://doi.org/10.1016/j.rse.2021.112466.

O'Keefe, J. 2019. "Phenology of Woody Species at Harvard Forest Since 1990."

Olsson, C., and A. M. Jönsson. 2014. "Process-Based Models Not Always Better Than Empirical Models for Simulating Budburst of Norway Spruce and Birch in Europe." *Global Change Biology* 20, no. 11: 3492–3507. https://doi.org/10.1111/gcb.12593.

Pastorello, G., C. Trotta, E. Canfora, et al. 2020. "The FLUXNET2015 Dataset and the ONEFlux Processing Pipeline for Eddy Covariance Data." *Scientific Data* 7, no. 1: 225. https://doi.org/10.1038/s41597-020-0534-3.

Piao, S., Q. Liu, A. Chen, et al. 2019. "Plant Phenology and Global Climate Change: Current Progresses and Challenges." *Global Change Biology* 25, no. 6: 1922–1940. https://doi.org/10.1111/gcb.14619.

Polgar, C. A., and R. B. Primack. 2011. "Leaf-Out Phenology of Temperate Woody Plants: From Trees to Ecosystems." *New Phytologist* 191, no. 4: 926–941. https://doi.org/10.1111/j.1469-8137.2011.03803.x.

Réaumur, R. A. 1735. "Observations du thermomètre faites à Paris pendant l'année, comparées avec celles qui ont été faites sous la ligne, à l'Isle de France, à Alger et quelques unes de nos iles de l'Amérique." *Mémoires l'Académie Royale des Sciences* 1735: 545–576.

Reich, B. J., and S. K. Ghosh. 2019. *Bayesian Statistical Methods*. New York: CRC Press.

Richardson, A. D., D. Y. Hollinger, D. B. Dail, J. T. Lee, J. W. Munger, and J. O'keefe. 2009. "Influence of Spring Phenology on Seasonal and Annual Carbon Balance in Two Contrasting New England Forests." *Tree Physiology* 29, no. 3: 321–331. https://doi.org/10.1093/treephys/tpn040.

Richardson, A. D., K. Hufkens, T. Milliman, et al. 2018. "Ecosystem Warming Extends Vegetation Activity but Heightens Vulnerability to Cold Temperatures." *Nature* 560, no. 7718: 368–371. https://doi.org/10.1038/s41586-018-0399-1.

Richardson, A. D., T. F. Keenan, M. Migliavacca, Y. Ryu, O. Sonnentag, and M. Toomey. 2013. "Climate Change, Phenology, and Phenological Control of Vegetation Feedbacks to the Climate System." *Agricultural and Forest Meteorology* 169: 156–173. https://doi.org/10.1016/j.agrformet.2012.09.012.

Richardson, A. D., and J. O'Keefe. 2009. "Phenological Differences Between Understory and Overstory." In *Phenology of Ecosystem Processes*, edited by A. Noormets, 87–117. New York: Springer. https://doi.org/10.1007/978-1-4419-0026-5_4.

Rousi, M., and J. Pusenius. 2005. "Variations in Phenology and Growth of European White Birch (*Betula pendula*) Clones." *Tree Physiology* 25, no. 2: 201–210. https://doi.org/10.1093/treephys/25.2.201.

Savage, J. A., and I. Chuine. 2021. "Coordination of Spring Vascular and Organ Phenology in Deciduous Angiosperms Growing in Seasonally Cold Climates." *New Phytologist* 230, no. 5: 1700–1715. https://doi.org/10.1111/nph.17289.

Seyednasrollah, B., A. M. Young, K. Hufkens, et al. 2019. "PhenoCam Dataset v2.0: Vegetation Phenology From Digital Camera Imagery, 2000–2018." *ORNL Distributed Active Archive Center*: 1674. https://doi.org/10.3334/ORNLDAAC/1674.

Templ, B., S. Fleck, and M. Templ. 2017. "Change of Plant Phenophases Explained by Survival Modeling." *International Journal of Biometeorology* 61, no. 5: 881–889. https://doi.org/10.1007/s00484-016-1267-z.

Templ, B., E. Koch, K. Bolmgren, et al. 2018. "Pan European Phenological Database (PEP725): A Single Point of Access for European Data." *International Journal of Biometeorology* 62: 1109–1113.

Thornton, M. M., R. Shrestha, Y. Wei, P. E. Thornton, S. Kao, and B. E. Wilson. 2020. "Daymet: Daily Surface Weather Data on a 1-Km Grid for North America, Version 4." https://doi.org/10.3334/ORNLDAAC/1840.

van der Schoot, C., and P. L. H. Rinne. 2011. "Dormancy Cycling at the Shoot Apical Meristem: Transitioning Between Self-Organization and Self-Arrest." *Plant Science* 180, no. 1: 120–131. https://doi.org/10.1016/j.plantsci.2010.08.009.

Vitasse, Y., S. Delzon, E. Dufrêne, et al. 2009. "Leaf Phenology Sensitivity to Temperature in European Trees: Do Within-Species Populations Exhibit Similar Responses?" *Agricultural and Forest Meteorology* 149, no. 5: 735–744. https://doi.org/10.1016/j.agrformet.2008.10.019.

Wang, H., J. Dai, J. Peñuelas, Q. Ge, Y. H. Fu, and C. Wu. 2022. "Winter Warming Offset One Half of the Spring Warming Effects on Leaf Unfolding." *Global Change Biology* 28, no. 20: 6033–6049. https://doi.org/10.1111/gcb.16358.

Wang, H., C. Wu, P. Ciais, et al. 2020. "Overestimation of the Effect of Climatic Warming on Spring Phenology due to Misrepresentation of Chilling." *Nature Communications* 11, no. 1: 4945. https://doi.org/10.1038/s41467-020-18743-8.

Wheeler, K. I., M. C. Dietze, D. LeBauer, et al. 2024. "Predicting Spring Phenology in Deciduous Broadleaf Forests: NEON Phenology Forecasting Community Challenge." *Agricultural and Forest Meteorology* 345: 109810. https://doi.org/10.1016/j.agrformet.2023.109810.

Wolkovich, E. M., J. Auerbach, C. J. Chamberlain, et al. 2021. "A Simple Explanation for Declining Temperature Sensitivity With Warming." *Global Change Biology* 27, no. 20: 4947–4949. https://doi.org/10.1111/gcb.15746.

Wolkovich, E. M., B. I. Cook, J. M. Allen, et al. 2012. "Warming Experiments Underpredict Plant Phenological Responses to Climate Change." *Nature* 485, no. 7399: 494–497. https://doi.org/10.1038/nature11014.

Xiang, Y., S. Gubian, B. Suomela, and J. Hoeng. 2013. "Generalized Simulated Annealing for Global Optimization: The GenSA Package." *R Journal* 5, no. 1: 13.

Zhang, H., I. Chuine, P. Regnier, P. Ciais, and W. Yuan. 2022. "Deciphering the Multiple Effects of Climate Warming on the Temporal Shift of Leaf Unfolding." *Nature Climate Change* 12, no. 2: 193–199. https://doi.org/10.1038/s41558-021-01261-w.

Zhang, H., S. Liu, P. Regnier, and W. Yuan. 2018. "New Insights on Plant Phenological Response to Temperature Revealed From Long-Term Widespread Observations in China." *Global Change Biology* 24, no. 5: 2066–2078. https://doi.org/10.1111/gcb.14002.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.

Appendix A

Methods and Data Sources

Models: We focus on five classic process-based spring phenology models that are the foundations of various other models: the Thermal Time (TT) model, the Parallel (PA) model, the Sequential (SQ) model, the Alternating (AT) model and the Unified (UN) model (Table A1; Supporting Information S1 for detailed model description). Models were implemented by the 'phenor' R package (Hufkens et al. 2018), and the generalised simulated annealing algorithm provided by the 'GenSA' R package (Xiang et al. 2013) was used to optimise these models, This approach has been widely used in many studies (e.g., Basler 2016; Chuine, Cour, and Rousseau 1998). As a comparison, we also considered a simple linear regression (LIN) model with preseason mean temperature as the sole predictor. We tested individual monthly mean temperatures from January to May as well as mean temperatures across months from January–March and March–May for the LIN model, and retained the best-fitting variant for comparison with the process-based models.

ΔAICc equations: The difference of AICc (ΔAICc) between two models ($\text{model}_1, \text{model}_2$) depends on their number of parameters (k_1, k_2) and residual sum of squares ($\text{RSS}_1, \text{RSS}_2$), or more commonly used $\text{RMSE}_1, \text{RMSE}_2$, in fitting the data.

$$\text{AICc} = 2 \times k + n \times \log\left(\frac{\text{RSS}}{n}\right) + \frac{2k(k+1)}{n-k-1} \qquad \text{(A1)}$$

$$\Delta\text{AICc} = 2\left(k_1 - k_2\right) + n \times \log\left(\frac{\text{RSS}_1}{\text{RSS}_2}\right) + \frac{2\text{k}_1\left(k_1+1\right)}{n-k_1-1} - \frac{2\text{k}_2\left(k_2+1\right)}{n-k_2-1} \qquad \text{(A2)}$$

$$= 2\left(k_1 - k_2\right) + 2\text{n} \times \log\left(\frac{\text{RMSE}_1}{\text{RMSE}_2}\right) + \frac{2\text{k}_1\left(k_1+1\right)}{n-k_1-1} - \frac{2\text{k}_2\left(k_2+1\right)}{\text{n}-\text{k}_2-1} \qquad \text{(A3)}$$

Data: There is considerable variability in the type, quality and spatial scale of phenological and temperature datasets available for fitting and testing phenological models. We bracketed this variability by evaluating spring phenology model fits using four different pairings of spring phenology and temperature data that are commonly used in the literature, from in situ measurements made by experts through coarse scale interpolated datasets (Table A2). First, the MCD12Q2 land surface phenology (LSP) dataset (500 m resolution; Friedl, Gray, and Sulla-Menashe 2019) and Daymet gridded temperature (1 km resolution; Thornton et al. 2020) were retrieved at eddy-covariance flux tower locations included in the FLUXNET2015 dataset (Pastorello et al. 2020) (Figure S1a) for 2001–2019. Our second combination retained the same LSP data but substituted temperatures directly measured at the flux tower sites (resulting in fewer site years than with Daymet due to missing data). The third combination used ground-observed phenology and temperature datasets created by experts at Harvard Forest (HF; O'Keefe 2019) in MA and Hubbard Brook Experimental Forest (HB; Bailey 2019) in NH, USA (Figure S1b). The fourth combination used the Pan European Phenological database (PEP725; Figure S1c; Templ et al. 2018) and the gridded E-OBS temperature dataset (Cornes et al. 2018). Species-specific phenological data in PEP725 were made by citizen scientists over decades. We chose six dominant species to conduct this analysis. The E-OBS dataset is a gridded, land-only observational climate dataset covering the European continent from 1950 to present with spatial resolutions of 0.1° and 0.25°. We obtained daily mean temperature at 0.1° (~10 km) spatial resolution from E-OBS at the PEP725 observational stations. The temperature data were then calibrated with a lapse rate using the elevation of PEP725 observational sites (Basler 2016; Zhang et al. 2022). Note that except for the PEP725 dataset, we did not investigate phenology data collected at many study regions. The selection of data sources and regions was based on the following criteria: (i) The MCD12Q2 + Daymet combination has been widely used in many previous studies to intercompare models since they are easy to access and have large spatial coverages (e.g., Melaas et al. 2016; Moon 2021). But Daymet is spatially interpolated, and it may not be as accurate as the ground-measured temperature. So, we selected MCD12Q2 + Daymet and MCD12Q2 + ground-measured temperatures at flux tower sites to investigate whether different conclusions can be drawn from these datasets. (ii) Following the same logic, we selected phenology data at HF and HB to investigate whether using field-observed phenology will

TABLE A1 | Models considered in the analysis.

Model	Process	Number of parameters	Citation
Thermal time (TT)	Thermal forcing, no chilling	3	Réaumur (1735)
Parallel (PA)	Chilling and thermal forcing accumulate simultaneously	9	Hänninen (1990); Kramer (1994)
Sequential (SQ)	Thermal forcing accumulates only after chilling is fulfilled	8	Hänninen (1990); Kramer (1994)
Alternating (AT)	Thermal forcing requirement follows an inverse exponential chilling–forcing function	5	Kramer (1994)
Unified (UN)	Unifies the TT, PA, SQ and AT models as special cases	9	Chuine (2000)
LIN	Linear regression against preseason mean temperature	1	Basler (2016)

TABLE A2 | Summary of phenological and temperature data sources and combinations used in this study. Note that most PEP725 sites recorded six species.

Phenology	Temperature	Number of sites	Number of site years	Location
MCD12Q2	Daymet	26	424	Forest eddy-covariance flux sites in North America
MCD12Q2	Flux temperature	22	156	
Ground-observed phenology	Ground-measured temperature	5	119	Harvard Forest, MA and Hubbard Brook Experimental Forest, NH
PEP725	E-OBS	1177	221,563	Europe

get different results compared to remotely sensed phenology. Although these two sites are spatially small and close, they may be the only places that have expert-observed publicly available long-term phenology data. (iii) Finally, although PEP725 + E-OBS has been used intensively in previous studies, conflicting results have been reported (e.g., Basler 2016; Wang et al. 2022; Zhang et al. 2022). Thus, we also selected this combination. We have also considered phenology data from USA National Phenology Network (USA-NPN) (Crimmins et al. 2022) and PhenoCam (Seyednasrollah et al. 2019), but both lack long-term site-specific data. Complete data processing details can be found in the Supporting Information S1, and all code is available online.