Hybrid machine learning approaches outperform mechanistic models of bloom timing in Eastern Redbud, Cercis canadensis

Tariq TSS Mohammad

tareqmohammed@arizona.edu

The University of Arizona https://orcid.org/0009-0009-1410-2482

Robert P Guralnick

Jorge AS Santiago-Blay

Theresa M Crimmins

Research Article

Keywords: Cercis canadensis, Eastern redbud, machine learning, phenological model

Posted Date: November 20th, 2025

DOI: https://doi.org/10.21203/rs.3.rs-7943611/v1

Version of Record: A version of this preprint was published at International Journal of Biometeorology on May 7th, 2026. See the published version at https://doi.org/10.1007/s00484-026-03197-2.

Hybrid machine learning approaches outperform mechanistic models of bloom timing in Eastern Redbud, *Cercis canadensis*

T. Mohammad[1,*], R.P. Guralnick[2], J.A. Santiago-Blay[3], T.M. Crimmins[4,5]

[1] Department of Mathematics, University of Arizona, Tucson, AZ, USA. ORCID: 0009-0009-1410-2482

[2] Florida Museum of Natural History, University of Florida, Gainesville, Florida, USA. ORCID: 0000-0001-6682-1504

[3] Department of Paleobiology, National Museum of Natural History, Washington, DC, USA and Penn State University, York, PA, USA. ORCID: 0000-0003-0219-5609

[4] USA National Phenology Network, Tucson, AZ, USA. ORCID: 0000-0001-9592-625X

[5] School of Natural Resources and the Environment, University of Arizona, Tucson, AZ, USA

* Corresponding author, tareqmohammed@arizona.edu

Author contributions

Conceptualization: TM, TMC, JASB;

Methodology: TM, TMC, RPG, JASB;

Formal analysis and investigation: TM, RPG;

Writing - original draft preparation: TM, TMC, RPG, JASB;

Writing - reviewing and editing: TM, TMC, RPG, JASB.

35
36
37
38

Abstract

Predicting the timing of flowering has utility both in climate change planning and practical applications. Eastern Redbud (*Cercis canadensis*), a charismatic spring-flowering tree found predominantly in the eastern USA, has long captured public attention, yet robust models of its bloom timing are lacking. Here, we aim to understand flowering onset dynamics of *C. canadensis*, a challenge given potential for local adaptation and plasticity across its range. We utilized a hybrid phenology modeling framework that integrates a mechanistic chill–heat framework with a machine-learning correction layer. We expected that this integrative model would better capture non-linear dynamics in flowering onset across its range. Using observations of bloom onset maintained by the USA National Phenology Network collected across the tree's range, we parameterized a process-based model incorporating chilling, forcing, and photoperiod cues, adjusted for latitude. This model predicted bloom dates with a mean absolute error of 7.3 days. Incorporating a machine-learning correction layer yielded a cross-validated mean error of 6.1 days, with improved representation of local anomalies and interannual variability. Out-of-sample validation using observations from herbarium and iNaturalist indicated the hybrid model retained predictive skill across contemporary and historical contexts. Our results also revealed that trees at higher latitudes require greater chilling and less forcing to flower than southern individuals. These results highlight the value of hybrid approaches that combine known extrinsic drivers linked to physiology with more flexible machine learning approaches, providing improved species-wide generalization. These analyses showcase both integrative modeling and integrative validation approaches to advance phenological forecasting.

Keywords: *Cercis canadensis*; Eastern redbud; machine learning; phenological model

Statements and Declarations

The authors declare no competing interests.

Introduction

Predicting the timing of flowering in socially and ecologically important species has been an active area of research for many decades. In previous centuries, the practice was largely motivated by economic reasons such as improving agricultural practices, improving plant breeds through hybridization and selection, and maximizing crop yield (e.g., Caprio 1966; Hopkins 1920; Hough 1866; Wood and Orel 2001; Zirkle 1935). More recently, the desire to predict when particular species of plants will bloom has also been undertaken to support tourism, to anticipate the start and peak of the allergy season, and to characterize the arrival of spring in real time (Chung et al. 2011; Hatzis et al. 2025; Zhang and Steiner 2022).

Eastern redbud, *Cercis canadensis* Linnaeus (Fabaceae), is a striking and highly visible harbinger of spring in the eastern United States. One of the earliest flowering trees in the region, it produces an abundance of pale to vibrant pink blossoms that emerge before leaf-out (Adkins et al. 2012; Dirr 2009). In the Cherokee language, Eastern redbud is known by several names that translate to "liar", in reference to the tree's tendency to flower before the last frost and falsely indicate the arrival of spring (Hamel and Chiltoskey 1975, 2002; Teuton et al. 2023). This highly visible springtime event has captivated humans' attention for centuries; in a letter to his daughter, Maria, dated May 8, 1791, Thomas Jefferson noted that "Red buds" were in flower in Philadelphia, Pennsylvania (Jefferson 1791). Due to its vibrant appearance, this native species is frequently used in landscaping and urban beautification projects. The tree's early-season blooms provide nectar for a wide range of pollinators, including hummingbirds, butterflies, and bees (Camilo et al. 2025; Koeser et al. 2022) and contribute to honey production (Lovell 1961).

Despite this species' ecological and societal importance, relatively little research has focused on developing predictive models for flowering in *C. canadensis*. Temperature, moisture, insolation, and winter chill have all been proposed as important variables shaping flowering time (Reader et al. 1974; Werle and Witcher 2022). However, the specific requirements to prompt flowering are largely unknown. The most replicable models available in the literature were published by Herms (2004), who established the quantity of spring warmth redbud trees must experience before flowering. Using observations collected in Michigan and Ohio, Herms (2004) calculated two distinct growing degree day threshold models for bloom in *C. canadensis*, reporting that trees in Michigan flowered after 177 growing degree days (GDD; base temperature 50°F, starting January 1), and trees in Ohio flowered after 191 growing degree days had been accumulated.

In this study, we aimed to establish a model to predict flowering time in *C. canadensis* across its range. Range-wide modeling can be challenging because many plants are known to exhibit local adaptation, suggesting that models established in one part of a species' range may not accurately predict flowering time in other parts of its range (Liang 2016, 2025). Multiple recent studies have indicated varying sensitivity to temperature among individual trees across the range of *C. canadensis*, which encompasses much of the eastern U.S., from New York to Florida and westward into Texas, Oklahoma, Kansas, and Nebraska (Dixon et al. 2024; Koeser et al. 2022), suggesting adaptation within the species to local conditions (Gallinat et al. 2025, Liang 2025). As such, models established using observations from only a portion of a species' range may be expected to exhibit decreasing predictive performance at locations progressively farther from the model origin.

Historically, the field of plant phenology research has relied on process-based (mechanistic) models that translate physiological cues into unified chill and forcing frameworks (Fishman et al., 1987; Chuine, 2000). These models remain the backbone of many contemporary phenology studies as they have proven to provide reliable phenological event forecasts. Yet, their fixed functional forms often fail to capture complex, non-linear interactions that can make range-wide modeling or modeling under rapid warming challenging, leading to systematic prediction error when it comes to extremes in budburst timing (Richardson et al., 2018). Recent cross-model comparisons show consistent under-estimation of particularly early or late leaf-out regions and years across six temperate species (Basler 2016). Moreover, large-scale evaluations indicate that seven of the twelve most common chilling models misrepresent the negative relationship between the accumulation of winter chill and spring heat requirements, which leads them to predict spring events (leaf-out and flowering) too early under long-term warming projections (Wang et al. 2020a).

Hybrid phenology modeling approaches that combine a mechanistic core with machine-learning correction layers are emerging as a promising alternative, retaining physiological realism while capturing residual climate signals that are otherwise unaccounted for. Similar formulations have demonstrated significant accuracy gains in diverse settings. Notably, a global soybean study fed process-model stage dates into a gradient-boosted tree ensemble and reduced error in predictions of flowering time by 36% (McCormick et al. 2021), and a model for predicting bloom in cherry trees, where a small neural block replaced the temperature-response kernel, outperformed both stand-alone process and machine learning baselines across Europe and east Asia (van Bree et al. 2025).

Using observations of *C. canadensis* collected across the plant's range by participants in *Nature's Notebook*, the USA National Phenology Network's plant and animal phenology observing platform, we

established a mechanistic sequential chill–heat to predict flowering in *C. canadensis*. To evaluate gains in predictive performance enabled by the addition of machine learning, we integrated the sequential model with a gradient-boosted tree ensemble using the same flowering observations. We tested model performance using both contemporary *Nature's Notebook* observations as well as out-of-sample contemporary and historical observations of redbud bloom transcribed from herbarium records and the popular iNaturalist platform. Finally, to test for evidence of local adaptation, we compared final model performance across the species' range to predictions generated using a GDD model established using observations from a single location in the extreme northern portion of *C. canadensis's* range.

Materials and methods

Cercis canadensis

Cercis canadensis (Eastern redbud) is a relatively small, deciduous tree native to eastern North America. It is commonly found in forests and moist woodlands and grows well in both full sun and partial shade (Dirr 2009; Gleason and Cronquist 1991; Robertson 1976). A member of the legume family (Fabaceae), *C. canadensis* is most easily recognized in early spring when it produces striking pink flowers, many of them directly on older stems and branches—a trait known as cauliflory (Figure 1). The flowers are zygomorphic (bilaterally symmetrical), resembling small butterflies, and are typical of the subfamily Faboideae (also known as Papilionoideae). Because of its ornamental appeal, Eastern redbud is widely cultivated in landscapes and planted internationally, with many horticultural cultivars available.

Fig. 1 Eastern redbud, Cercis canadensis. A. Tree in flower eastern Tennessee, USA. Photo credit: M.

Gillespie. B. Close-up of open flowers, taken in Howard County, Maryland, USA. Photo credit: H. Lowe Metzman

USA-NPN phenological data preparation

Nature's Notebook (www.naturesnotebook.org), hosted by the USA National Phenology Network (USA-NPN), is a leading effort to document the timing of seasonal activity – phenology – in plants and animals in the U.S. The *Nature's Notebook* platform consists of standardized observation protocols, a mobile app for data collection, data visualization and download tools, and an online interface for submitting and accessing observations. In 2021, the USA-NPN established a formal effort to collect information about redbud phenology through *Nature's Notebook*. The USA-NPN's Redbud Phenology Project campaign is widely advertised through emails to *Nature's Notebook* participants and other volunteer networks, in-person presentations, and through scientific publications (Brewer and Santiago-Blay 2022a, b; Posthumus et al. 2021; Tighe et al. 2024). USA-NPN staff and campaign leads hosted a kick-off webinar each winter (December or January) to recruit new participants and introduce the basics of observing. Campaign participants are emailed approximately every six weeks during the growing season with project updates, data summaries, other news related to redbuds, and the human touch of heartily felt encouragement. Since the initiation of the campaign, observation of redbuds in *Nature's Notebook* has more than doubled, with over 200 sites reporting each year since 2022 (Crimmins et al. 2024). To date, several hundred volunteers have contributed more than 1,000 observations across the U.S., with most observations concentrated in eastern states (Crimmins et al. 2024).

The *Nature's Notebook* phenology observation protocols are "status" protocols, meaning that participants report on the status of life cycle stages for an individual plant or animal species each time they make an observation (Denny et al. 2014). Participants are encouraged to observe frequently throughout the growing season to capture the onset of events such as flowering with as much temporal precision as possible. Each plant phenology observation is composed of "yes" or "no" responses to a series of questions pertaining to the status of a plant's leaves, flowers, and fruits.

We downloaded the USA-NPN "individual phenometrics" data for *C. canadensis* for the period encompassing Mar 21, 2011– May 6, 2025 (downloaded May 23, 2025; USA National Phenology Network, 2025). Individual phenometrics are derived estimates of phenophase onset and end dates within a period of interest. To minimize erroneous records, we retained only the first annual observation of "open flowers" for each tree in each year. Further, we retained records where a "yes" record for "open flowers" was preceded by a "no" report in the previous 14 days, an approach implemented in other recent

studies using USA-NPN observations (e.g., Crimmins et al. 2017, Liang 2025). To ensure biological plausibility, we discarded any first "yes" records occurring after day of year 181 (≈ 30 June). This cutoff is quite conservative, extending well beyond the documented bloom window for C. canadensis, which typically flowers in early spring (Mar–May; Brakie 2010). We employed a global 3×MAD (median absolute deviation) screen, a robust dispersion measure centered on the median, to remove extreme global outliers, a standard practice when volunteer-collected data (Belitz et al., 2022). Unlike standard deviation-based cutoffs, which can be distorted by a handful of large values, MAD resists the influence of extreme values and provides a stable threshold even when the data contain gross errors (Leys et al. 2013). We then performed an additional filter within each 5° latitude band, discarding records that fell outside the 2.5th–97.5th percentile range of observations for that band. This stratified trimming approach preserves a broad latitudinal continuum of records while eliminating local anomalies; in contrast, a single global percentile cutoff would risk discarding valid early southern or late northern events simply because they occupy the tails of the overall distribution (Morin et al. 2009). Finally, we only retained records submitted by observers bearing an email account with a ".gov" or ".edu" extension, surmising that observers with these email addresses were more likely to be undertaking phenology observing as a part of their jobs, and therefore may yield observations with greater accuracy. After all quality controls were implemented, we retained 179 discrete instances of flowering onset spanning 25°-50°N.

For each observation, we retrieved an 800 m resolution daily minimum temperature, maximum temperature, and precipitation time series from the PRISM gridded climate dataset, covering the full temporal period of our phenological records and our spatial extent in the eastern USA (PRISM Climate Group, 2024). Temperature values were converted from °F to °C.

iNaturalist and herbarium phenological data

To test the performance of our model to predict flowering time in both recent years and several decades in the past, we compiled records of *C. canadensis* in bloom from both herbarium specimens and the iNaturalist platform. For herbarium records, we focused on records from 1980 to present to maximize accuracy and observation quality for use in backcasting -- assessing the model's performance at predicting onset dates in the past -- at a reasonable resolution. First, we searched iDigBio, GBIF and Symbiota plant-based portals for "Cercis canadensis," including fuzzy matching for synonyms or for subspecies. We limited records to those between 1980 and present with images. Co-author RPG annotated images to record those that were "early flowering." We defined "early flowering" specimen records as those exhibiting a mix of open flowers and flower buds but no evidence of unfolded leaves. Next, we took screenshots of specimen labels of the "early flowering" images and used ChatGPT4 to assemble metadata

from those labels including date, and georeferences, if those data were not already available. Co-author RPG spotchecked all records for quality and further validated a subset of the more complicated georeferences, e.g. with multiple directional offsets, and all dates, to assure accuracy. In particular, he spotchecked 10% of all records and for the most difficult ones (~10%), he checked accuracy and manually adjusted decimal latitude and longitude if there were problems.

We downloaded all images with "flower present" annotations from iNaturalist contributed in the years 2018-2024. For observations contributed in the years 2010-2017, co-author RPG directly examined all research grade Eastern Redbud images that had public locations, georeferences and dates metadata and annotated those showing early flowering redbuds. He then assembled a spreadsheet capturing record identifiers, date and location metadata, including latitude and longitude, for early flowering annotations. Most of these records had directly reported GPS coordinates; in cases where those were missing we could often georeference them based on available information in the record. In total, we used 102 records from herbaria and iNaturalist for backcasting, with the iNaturalist data better representing the 2017-2024 period and herbarium better representing 1980-mid 2010s.

Records of "open flowers" contributed to iNaturalist and herbarium records generally reflect a slightly different segment of the seasonal flowering pattern than those captured through the status monitoring approach embedded in *Nature's Notebook*. Herbarium specimens and iNaturalist photos are typically collected once flowers have at least begun to open; records used in this study best reflect "early flowering;" in contrast, the individual phenometrics derived from *Nature's Notebook* observations indicate a window in time within which flowers transitions from not yet open to open. As such, we expect our model, which is tuned to predict the date flowers first open, to exhibit systematic bias when applied to herbarium and iNaturalist records. As well, herbarium specimens and iNaturalist images almost always represent a portion of a plant. It is possible that another, undocumented part of the tree exhibited evidence of later phenology that is not reflected in the specimen or iNaturalist record. Due to these issues, the records and dates of "early flowering" derived from iNaturalist images and herbarium records on parts of plants are likely more uncertain compared to observations made on entire plants and contributed to *Nature's Notebook*.

Model construction

Let (ϕ) represent the latitude at a particular site, and let ($\underline{\phi}$) describe the mean latitude of the training set. To predict flowering time in *C. canadensis*, we implemented a Utah-style chill accumulator, which

assigns an hourly chill weight in line with the temperature-bin scores established in the original peach-dormancy work of Richardson et al. (1974)[1]. We chose the Utah model as our chill formulation because it has previously been widely applied in large-scale phenology analyses and offers robust performance with citizen-science datasets (Luedeling 2012). Chill units accumulate starting on an onset day $D(\phi)$ until the completion of a Chill requirement $C(\phi)$. While many models assume a fixed chill start date, empirical evidence shows that autumn leaf senescence and dormancy onset advance systematically with latitude (Vitasse et al. 2009; Fu et al. 2014; Delpierre et al. 2016; Zohner et al. 2017). Chill fulfillment alone, however, offers an incomplete physiological description, as controlled-environment studies show that buds remain dormant until day-length cues signal favorable spring conditions (Laube et al. 2014; Way and Montgomery 2015). Accordingly, we implemented a latitude-variable photoperiod gate, restricting bud development until the day length met the critical threshold $P(\phi)$. After the photoperiod gate opens, growing-degree days (GDD) above a variable base temperature accumulate towards a forcing threshold $F(\phi)$, which, once satisfied, results in flowering (Basler 2016; Chuine 2000). To capture geographic adaptation across the species' broad latitudinal range, we let every primary threshold drift linearly with latitude (Δ_{lat}), yielding the following latitude-dependent formulations:

$$\text{Latitude offset: } \Delta_{\text{lat}} = \phi - \underline{\phi}$$

$$\text{Chill Requirement: } C(\phi) = C_0 + C_s\Delta_{\text{lat}}$$

$$\text{Forcing requirement: } F(\phi) = F_0 + F_s\Delta_{\text{lat}}$$

$$\text{Photoperiod limit: } P(\phi) = P_0 + P_s\Delta_{\text{lat}}$$

$$\text{Chill-onset day: } D(\phi) = D_0 + D_s\Delta_{\text{lat}}$$

In addition to these base formulations, three additional mechanisms further refine the model. First, the forcing requirement is adjusted dynamically with accumulated chill, consistent with evidence that insufficient chilling increases heat requirements, while abundant chill reduces them (Hänninen 1990; Wang et al. 2020b). When winter chill does not meet the chill requirement, the model increases the forcing demand by a deficit sensitivity factor (γ_{def}). In contrast, excess chill reduces the forcing demand by an equivalent excess sensitivity factor (γ_{exc}). This relationship can be expressed as:

[1] Chill Score Assignments: (< 1.4 °C = 0; 1.4–2.4 °C = 0.5; 2.4–9.1 °C = 1; 9.1–12.4 °C = 0.5; 12.4–15.9 °C = 0; 15.9–18 °C = –0.5; > 18 °C = –1).

$$F^* = F + \frac{1}{2}(\gamma_{\text{def}} + \gamma_{\text{exc}})(C - C_a) + \frac{1}{2}(\gamma_{\text{def}} - \gamma_{\text{exc}})|C - C_a|$$

Where F is the latitude adjusted forcing requirement (in growing degree days, GDD), F* is the chill-adjusted forcing requirement, C is the chill requirement (chill units, CU), and C_a is the accumulated chill at the site in a given year. The parameters γ_{def} and γ_{exc} control the strength of compensation under chill variability. When chill accumulation falls short of the requirement ($C_a < C$), the model increases the forcing requirement by a factor of γ_{def} per unit chill deficit. Conversely, when chill accumulation exceeds the requirement ($C_a > C$), the model decreases the forcing requirement by a factor of γ_{exc} per unit chill surplus. This formulation ensures the forcing threshold dynamically compensates for interannual variability in chill conditions, allowing the model to reproduce documented plasticity across different winter conditions. These thresholds allow the model to capture population-specific differences across the species' broad range. In the warmest parts of this range, *C. canadensis* often reaches anthesis despite chronically low winter chill. We therefore implement a latitude-conditioned override mechanism. For sites below ($\phi \leq 35°N$), we suppress latitude slopes on chill/photoperiod/onset parameters, effectively setting:

$$C_s = P_s = D_s = 0.$$

We then apply an early spring override deadline (O_{min}), where

$$O_{\min}(\phi) = \max(O_{\min} + O_s\Delta_{\text{lat}}, 15).$$

During simulation, if on any given day of the year (d) in the bloom year the accumulated chill (C_a) has not yet met the chill requirement (C) and $d \geq O_{\min}(\phi)$, we force the fulfillment of the chilling requirement. Following the forced fulfillment, the photoperiod gate and heat accumulation proceed normally towards the dynamic forcing requirement (F^*). The second value of the max term (15, DOY) represents a conservative lower bound we have selected to prevent unrealistically early overrides at far-southern sites. This process functionally bifurcates the population into a standard (chill-limited) cohort and an override (heat-limited) southern cohort, in line with empirical evidence that trees belonging to subtropical regions can bypass classical chill thresholds via alternative hormonal pathways (Luedeling and Brown 2011; Rodríguez et al. 2019). There is also sufficient evidence to indicate a phenological breakpoint in first-flowering response across North America, with populations south of this transition exhibiting distinct behavior under limited chilling conditions (Perrier et al., 2025).

Finally, bud responsiveness to temperature is not constant, but can acclimate modestly to the climatic context of the season (Basler 2016; Fu et al. 2015; Hänninen 1990). To reflect this, we allowed the forcing base temperature to vary with background spring conditions, such that the effective threshold for heat accumulation (T_b^*) would be expressed

$$T_b^* = T_b + T_w T_{\text{spring}}$$

where T_b is the threshold temperature used to compute each day's contribution to GDD, T_{spring} is the mean daily temperature during the pre-bloom season of each year (January–March for most sites), ensuring that the adjustment reflects only the climatic context leading up to flowering, and T_w is a slope parameter that controls the degree of adjustment. This feature helps the model formally accommodate plastic acclimation in thermal response, improving its realism across diverse sites and years.

To identify the best combination of the 14 mechanistic parameters, we employed a two-stage metaheuristic workflow that explored the parameter space within biologically sensible bounds (see Table 1 for parameter bounds) to minimize the mean absolute error (MAE) between predicted and observed flowering dates. We optimized the 14 parameters with Differential Evolution (DE; Storn and Price 1997), which maintains a population of candidate parameter vectors, mutates them by scaled vector differences, and replaces old candidates with better-scoring trial vectors in each generation. Search ranges were restricted to biologically plausible limits; hence, we employed a population of 12 vectors sufficient to cover the space without the need for unnecessary evaluations (≈ 0.8 × the parameter dimension; Das and Suganthan 2011). Our algorithm used a mutation factor (F) 2 drawn uniformly from the interval [0.5, 1.0] to ensure a mixture of larger exploratory steps and smaller exploitative steps during optimization. Our DE implementation used the canonical binomial recombination rate of 0.7 (Storn and Price 1997) and was run for a total of 140 generations, after which improvements in MAE plateaued. The best scoring vector was then passed to a Nelder–Mead simplex optimizer (originally described in Nelder and Mead 1965), which iterated until both the largest parameter change and the MAE change fell below 10^{-4}, or until 500 iterations had elapsed. Differential Evolution is well established in phenological model calibration when exact parameter values are unknown, for example, Taylor et al. (2019) calibrated temperate-tree phenology models with DE, and Kimura et al. (2020) used DE to fit a bud-burst model for *Camellia sinensis* (tea plant).

The resulting parameterized mechanistic model served as the baseline for a second, machine-learning correction layer, which builds on this baseline by learning the subtle site-specific and short-term thermal

[2] F is a positive scalar that scales the difference between two randomly chosen population members.

cues that the primary mechanistic model may not have picked up on. We implemented this second stage correction with a gradient boosting tree ensemble (LightGBM, an efficient, histogram-based implementation of gradient boosting) chosen for its ability to capture non-linear interactions (Friedman 2001; Ke et al. 2017). The gradient-boosted tree learner takes the predictions produced by the mechanistic model together with short-term temperature statistics, recent chill sums, day length, and latitude inputs, allowing the learner to correct remaining biases and capture nonlinear signals that the fixed equations may have missed. We represented each observation by nine code-defined predictors: the mechanistic model's DOY forecast, the 30-day mean and 30-day variance of daily temperature, cumulative chill over the previous 7 and 30 days, and the mean pre-spring temperature (tspr). Hyperparameters were held fixed following a preliminary Bayesian sweep with Optuna, an automatic, tree-search hyperparameter optimizer (Akiba et al., 2019). This tuning process identified a learning rate of 0.05, which controls the contribution of each successive tree to the ensemble, ensuring the model learns gradually enough to guard against overfitting while converging efficiently. The maximum number of leaves per tree was set to the default 31, with a minimum of 20 observations per leaf. We allowed the model 10,000 boosting rounds, providing ample capacity for the model to converge. To prevent unnecessary computation and overfitting, we invoked automatic halting after 50 rounds without improvement. Together, these settings ensure the model converges reliably while maintaining strong generalizability across sites.

Model evaluation using Nature's Notebook observations

To assess model performance, we implemented a five-fold cross-validation framework. Our 179 observations were partitioned into five mutually exclusive folds using a site-blocking scheme, such that records from the same geographic location were never split across training and validation sets. This reduces the risk of inflated accuracy due to spatial autocorrelation. For each fold, the model was trained on four partitions and evaluated using the fifth partition. Predictive accuracy was quantified using the mean absolute error (MAE) between predicted and observed day of year of first flowering, and final performance was summarized as the average MAE across all five folds.

We also report SHAP (SHapley Additive exPlanations) values (Lundberg and Lee 2017) to interpret the behavior of the machine-learning correction layer. SHAP provides a framework that breaks each prediction into a baseline value and the contribution of individual predictors. This allows us to quantify how much each variable influences the bloom-date forecast and ensures interpretability even in non-linear models such as gradient boosting.

As a baseline for comparison, we also evaluated the growing degree threshold published by Herms (2004). Implementing this model at observation sites across *C. canadensis's* range allowed us to directly evaluate whether this species appears to exhibit local adaptation or plasticity. We applied the fixed 177 GDD threshold across our full dataset, which spans the whole native geographic range of Cercis canadensis and contains more than a decade of observations. We use the same 800 m resolution daily climate time series employed in calibrating our hybrid model.

Model evaluation using herbarium and iNaturalist records

We evaluated generalization to an independent dataset of historical and recent observations of early flowering in Eastern Redbud by then reapplying the final hybrid model without refitting parameters to compiled from herbarium and iNaturalist records. Records were pre-conditioned using the same filtering and processing steps described above. For each site–year, we reconstructed climate sequences for the prior winter and year of bloom, and ran the unchanged mechanistic model to obtain a baseline prediction. We then computed the same ML feature set used during training, namely, the mechanistic prediction (*mech_pred*), the 30-day mean and variance of daily temperature (*temp_mean_30d*, *temp_var_30d*), 7-day and 30-day cumulative chill (*chill_7d*, *chill_30d*), and mean pre-spring temperature (*tspr*).

Results

After quality-control filters were applied, 179 unique site × year flowering observations remained for analysis (Fig. 1). The retained observations were broadly distributed across the documented distribution of *C. canadensis*, with the greatest density of records in the northeastern portion of the range. Of the 152 sites, 33 contributed records across multiple years, providing temporal replication that supported the calibration dataset.

Prediction performance of the GDD threshold model

Application of Herms' (2004) GDD threshold across our dataset yielded predicted bloom dates that were consistently later than observed phenological events. The threshold produced a mean residual of –38.2 days (Observed - Predicted) and a median residual of –37 days, indicating that the fixed 177 GDD requirement systematically overestimates the thermal forcing needed for first bloom in *C. canadensis*. There is a weak pattern of smaller residuals occurring at observation sites closer to the model's origin in Midland, MI (Fig. 2a).

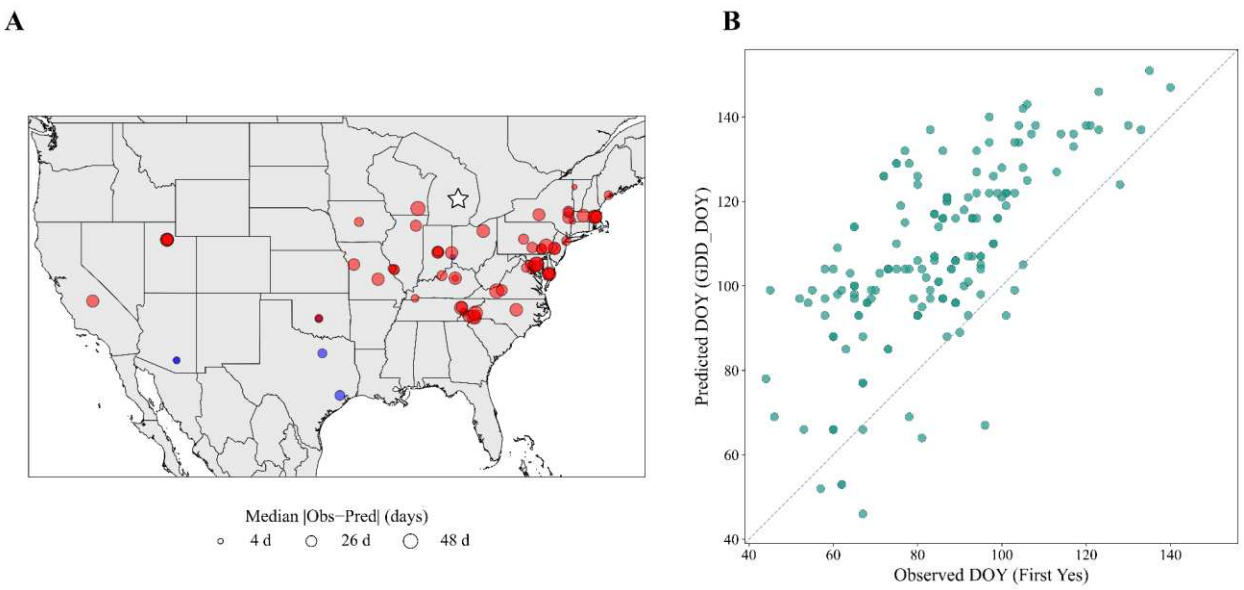

Fig. 2 Performance of Herms' (2004) 177 GDD threshold applied to *C. canadensis* flowering records. a) Observed versus predicted day of year (DOY) of first bloom across 206 individual site × year records. b) Distribution of residuals showing that the threshold consistently predicts bloom later than observed, with most residuals clustered around –30 to –40 days

Mechanistic model performance

Optimization of the mechanistic model yielded biologically consistent parameter estimates (Table 1). The fitted chill onset (D0) (DOY 286≈ mid-October) and onset-lat slope (Os ≈ +1.26 days °N^{-1}) indicate that chilling is initiated relatively late in the calendar year, with slightly later onset toward northern latitudes. The baseline chill requirement (C_0 ≈ 663 CU) increased with latitude (Cs_1 ≈ +1.9 CU °N^{-1}), while the baseline forcing requirement (F_0 ≈ 433 GDD) decreased with latitude (F_s ≈ –2.8 GDD °N^{-1}), capturing compensation between chilling and forcing across the geographic gradient. Photoperiod parameters showed a critical threshold of ~11.1 h (Ph_0), with a modest positive slope (P_s ≈ 0.06 h °N^{-1}), implying a limited gating role.

Table 1. Parameter estimates for best-fitting mechanistic model predicting bloom in *C. canadensis*

Symbol	Description	Estimate	Lower	Upper	Units/notes
D_0	Chill-onset DOY	286.428	240	350	DOY
D_s	Chill-onset latitudinal slope	1.257	-3	3	Days per °lat

C_0	Chill requirement	662.595	400	1200	Chill Units (CU)
C_s	Chill requirement latitudinal slope	1.879	0	5	CU per °lat
P_0	Critical photoperiod	11.067	9	15	Hours
P_s	Photoperiod latitudinal slope	0.062	-0.1	0.1	Hours per °lat
F_0	Forcing requirement	432.719	50	500	Growing Degree Days (GDD)
F_s	Forcing requirement latitudinal slope	-2.801	-3	3	GDD per °lat
γ_{def}	Chill-deficit sensitivity	0.703	0	1	unitless
γ_{exc}	Chill-excess sensitivity	0.982	0	1	unitless
O_{min}	Baseline earliest spring override DOY	43.402	15	120	DOY
O_s	Latitude slope on the override date (O_{min})	-1.850	-8	8	days per °lat
T_b	Base temperature for forcing	1.160	1	5	°C
T_w	Spring-temp slope adjusting T_b	-0.835	-1	1	°C change in T_b per °C spring temp

Optimization produced a mechanistic fit with MAE = 7.25 days (Fig. 2A). Residuals were approximately centered, indicating that the model neither consistently under- or over-predicted the date of bloom (mean

= 2.05 ± 9.25 days; Fig. 2B). Latitude specific trending (Fig. 2C) indicates the model captures a modest increase in chill requirements (+1.9 CU deg^{-1}) and a compensatory decrease in forcing requirements (–2.8 GDD deg^{-1}) at higher latitudes.

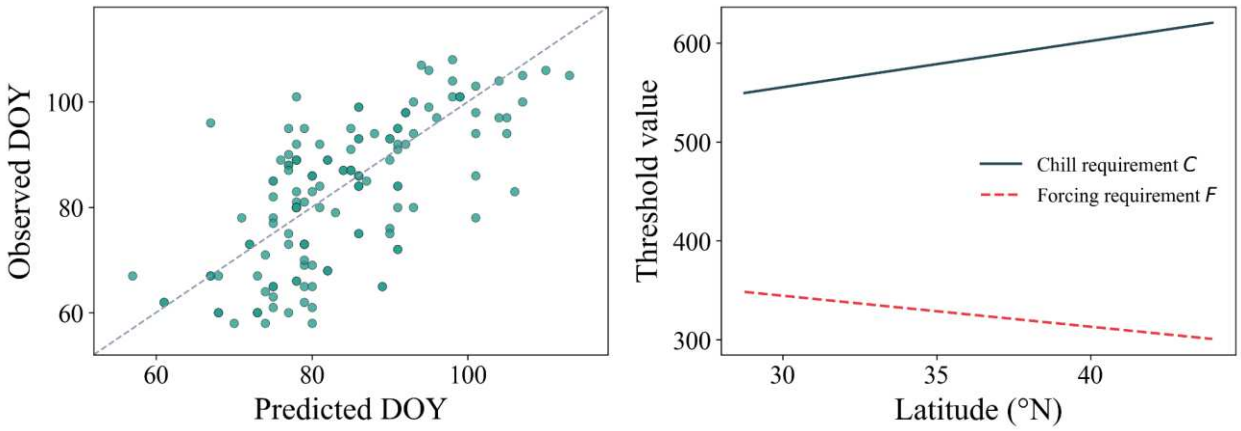

Figure 2. a) Scatterplot of observed versus predicted flowering dates for the 179 calibration records. The dashed line denotes a one-to-one fit. b) Latitude dependence of the calibrated thresholds: the chill requirement, C (teal), increases by approximately two chill units per degree north, whereas the forcing requirement F (dashed orange) decreases by approximately three growing degree days per degree north.

The mechanistic model's errors were essentially unbiased in space and only weakly biased in time. A plot of model residuals shows no discernible relationship between the magnitude of residuals and latitude (r = 0.01, p = 0.92), with a cloud of points symmetrically centered across the full range (Fig. 3a), revealing no evidence of systematic latitudinal bias. A marginal upward drift in residuals through time is evident (r = 0.15, p = 0.04, Fig. 3b); the pattern coincides with a marked rise in observation volume in the later years of the record.

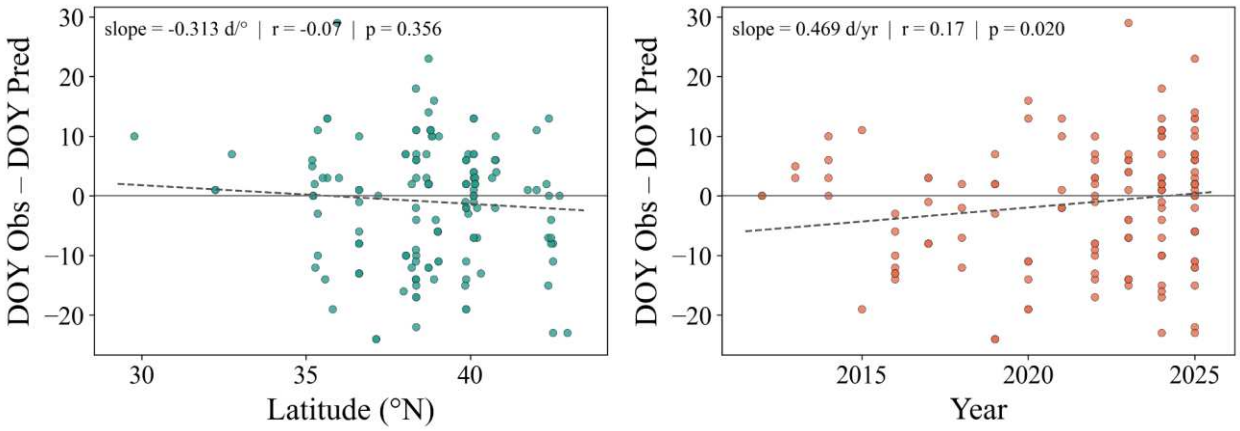

Fig. 3 a) Residuals plotted against latitude. The fitted dashed line is the least-squares regression; b) Residuals versus year for the temporal record

Machine learning-enabled model improvements

Application of the LightGBM correction layer improved the predictive accuracy relative to the standalone mechanistic model by ≈1.12 days (from 7.26 to 6.14 days MAE, ~15.4% reduction). Nested 5-fold site-blocked cross-validation returned fold MAEs of 5.32, 4.53, 5.30, 6.92, and 8.65 days, yielding a final honest cross-validated mean MAE of 6.14 ± 1.48 days. Model interpretability (global mean |SHAP| from the full data refit) indicates that the mechanistic model (mech_pred) remains the dominant explanatory signal, with additional refinement coming from recent temperature mean (30-day), chill accumulation (30- and 7-day windows), spring temperature conditions (tspr), and 30-day temperature variance (Fig 4).

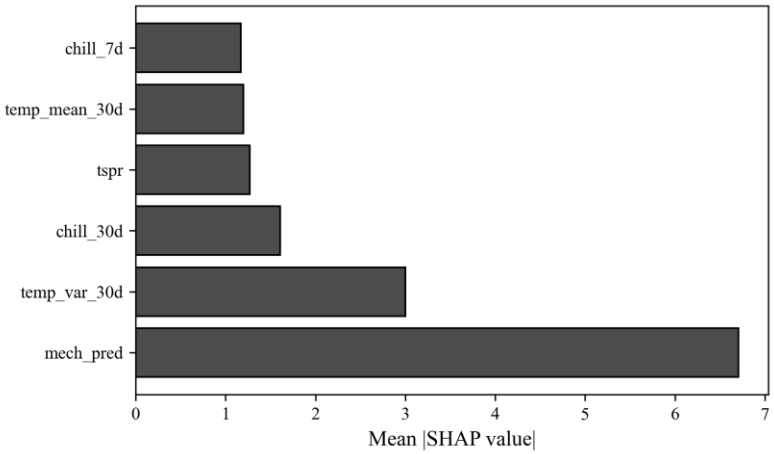

Fig. 4 Post LightGBM diagnostics. Global Feature Importance SHAP metrics, showing that mech_pred remains the primary predictor, while short-term climactic variables contribute additional refinements

Machine learning-enhanced model performance with out-of-sample historical observations

We evaluated the final hybrid model using 102 flowering observations compiled from herbarium records and iNaturalist, restricted to the domain observed during training to avoid extrapolation beyond the climatic ranges over which the model's parameters were optimized. No parameters were refitted, and no spatial or bias correction measures were applied. Across all sites, the standalone mechanistic model yielded an MAE of 19.58 days. The hybrid model, incorporating the machine learning layer, produced a MAE of 19.18 days with this same set of observations. Both models showed a consistent positive offset, with flowering predicted approximately 2.5 weeks earlier than reported. This offset was uniform rather than geographically structured, as residuals did not show any significant correlation with latitude (r= 0.13, p= 0.2). This systematic bias was expected given that herbarium and iNaturalist records capture "early flowering" rather than onset.

Discussion

Cercis canadensis, long viewed as a harbinger of spring, is an ecologically and economically important plant. In this study, we endeavored to develop a model to predict flowering in *C. canadensis*, a highly visible springtime phenological event. We explored whether the addition of machine learning enhancements led to improvements in prediction performance over a traditional mechanistic model; our findings have implications for predicting the timing of phenological transitions in a wide range of plant taxa.

Mechanistic models are enhanced by a machine learning layer

Using observations of flowering onset collected across the native range of *C. canadensis* in the U.S., we were able to establish a mechanistic model with strong performance and minimal spatial bias. On average, the mechanistic model, which incorporated the traditional variables of chill, forcing, and daylength, predicted flowering within one week of the date the event was reported by observers. Actual model error is likely less; the "first open flower" dates reported by *Nature's Notebook* observers may be as many as 13 days later than when the event actually occurred. While *Nature's Notebook* participants are encouraged to observe frequently during the spring season when plants are undergoing rapid change, observation frequency can be highly variable. We filtered our dataset to retain reports of "open flowers" preceded by a "no" report within the previous 14 days, meaning that the flowers could have actually opened anywhere within that window of time.

The addition of the machine-learning layer improved model performance by allowing the hybrid framework to capture non-linear context-dependent deviations, which may not have been explicitly captured by the mechanistic portion of the model. While the mechanistic model encodes the physiological processes that govern flowering, such as chilling accumulation, photoperiod gating, and thermal forcing, it simplifies these processes and assumes a consistent parameter-response relationship across space. The ML layer corrects this by learning residual patterns that arise from spatially variable factors, such as interactions between climatic variables. In effect, the mechanistic core anchors predictions in physiological realism, while the ML component adjusts predictions based on empirical signals arising in the data. Implementing the model to predict out-of-sample flowering dates revealed a similar pattern, where the mechanistic model performed well, and the addition of the machine learning layer further improved predictive performance. Overall, predictions made for the historical observations derived from herbarium records exhibited systematically larger errors that were biased to predict flowering earlier than it was reported. This systematic bias results from a mismatch between the plant developmental stage used in model prediction and evaluation. Whereas the observations used to construct the model are dates that flowers were first reported as open, the herbarium records best reflect plants in an early stage of flowering, likely to be several days or weeks later than when the plant first exhibited open flowers. When this fact is considered, the model is likely performing well at predicting the timing of flowering onset.

Some approaches using historical or current incidental data estimate flowering onset timing by aggregating records within a spatial region and then utilizing models that are appropriate for capturing the first events, such as the Weibull estimator (Belitz et al. 2020; Pearse et al. 2017). Capturing onset events using these approaches requires significant data densities in a year and region for accurate estimation, which is often a challenge given data sparsity (Belitz et al. 2022) and even then, uncertainty remains. Given this, we preferred using more site-specific data, while expecting a systematic bias toward early dates in modeled onset timing compared to the early-flowering empirical data. Overall, the backcast demonstrates that while the hybrid model marginally improves precision, the ML layer cannot substantially correct for systematic temporal offsets in independent data. This arises from both the training objective and the way LightGBM functions. The LightGBM layer was trained to model residual variation around the mechanistic predictions using short-term climatic features such as temperature means, variance, and chill accumulation. In line with gradient-boosting frameworks, it builds a series of decision trees that sequentially reduce the remaining error from previous trees, which leads to fine-scale corrections based on the features specified rather than inferring a baseline shift in flowering required to offset the 2.5 week discrepancy.

Because the backcast applied the model without retraining and did not include, importantly, any bias corrective measures, the ML layer operated within the scope of patterns it had previously learned, refining mechanistic timing based on within-season climatic anomalies, not recalibrating the overall baseline, hence the slight improvement. This behavior highlights that the framework is not a black box capable of arbitrarily fixing model deficiencies, but a structured layer that operates strictly within the relationships it is trained to capture. It learns how to recognize subtle, repeated departures from the mechanistic baseline, such as how climatic variables systematically nudge flowering earlier or later within a season. This makes these implementations powerful at capturing non-linear interactions, yet still, fundamentally anchored to the training domain provided by the mechanistic stage. It may be possible in the future to calibrate joint models that can incorporate incidental and more semi-structured survey data into a full framework that may further capture anomalies. This may be particularly valuable given that we had to limit our out-of-calibration data to the same latitudinal extents as our calibrated model, suggesting that more work is needed to capture Eastern Redbud flowering at the southern edges of its distribution.

Evidence of geographic adaptation

The results of our modeling efforts suggest local adaptation in *C. canadensis*'s sensitivity to temperature. The models we established in this study indicated that individuals at higher latitudes required exposure to slightly greater chill as well as slightly less warmth to induce flowering than individuals at lower latitudes. This pattern is consistent with a co-gradient, where individual trees located in colder reaches of the species' range flower later than individuals in warmer locations under otherwise constant conditions (Liang 2016). A co-gradient response can emerge when individuals in colder regions require exposure to greater chill than individuals in warmer regions to trigger leaf-out or bloom. Evidence of a co-gradient pattern has been documented in several early-spring flowering trees, including black cherry (Barnett and Farmer 1980), silver maple (Ashby et al. 1992), and paper birch (Hawkins and Dhar 2012).

The findings emerging from our implementation of Herms' (2004) simple GDD threshold further support the conclusion that *C. canadensis* exhibits significant range-wide variation in its physiological responsiveness to climate. In particular, the fact that the threshold, established using observations from one location, exhibited poor predictive performance in other portions of the species' range suggests the presence of adaptation or plasticity. As well, the over-estimation of the amount of warmth required to initiate flowering at other locations in the tree's range, which are all farther south and overall warmer, suggests that *C. canadensis* individuals in warmer environments require less warmth to initiate flowering, consistent with a co-gradient pattern (Liang 2016). Our finding of evidence of different local

physiological responsiveness in *C. canadensis* is consistent with recent studies that have suggested this (Gallinat et al. 2025; Liang 2025). These findings suggest that the ML-enhanced model, which learns local patterns such as variable sensitivity to the factors shaping flowering, is likely to demonstrate the strongest predictive performance among model types evaluated.

Strengths and weaknesses of volunteer-contributed data

The recent, rapid growth of citizen and community science platforms such as *Nature's Notebook* and iNaturalist has resulted in a wealth of observations contributed across large geographic regions. These data resources are rapidly opening doors to establishing models to predict phenological transitions such as flowering in a wide range of species. Because the data are available across large geographic regions, they are also unlocking new insights, such as the extent to which local adaptation exists.

Involving non-experts in data collection can introduce concerns about observation accuracy. Several studies have demonstrated strong performance among *Nature's Notebook* observers to accurately assess phenophase status (Feldman et al. 2018; Fuccillo et al. 2015). Similarly, iNaturalist observations have proven trustworthy and well-suited for various modeling applications (Dinnage et al. 2025; Larsen et al. 2022). Redbud is especially well-suited for observing by volunteers, common on the landscape and bearing plant structures and phenophases that are easy to identify and assess. The flowers are highly visible: they are large compared to many other trees, brightly colored, and emerge before tree canopies fill with leaves (Fig. 1). When flowers have transitioned from "not yet open" to "open" is clearly apparent. For these reasons, volunteer observers have long been trusted to track redbud phenology. A statewide effort based in North Carolina asked volunteers to track the timing of bloom in *C. canadensis* in the 1970s (Hopp 1974). As well, *Cercis chinensis*, a closely related species, has been widely observed in Asia, presumably due to the tree's highly visible and attractive flowers. Observations of flowering in *C. chinensis* comprise some of the longest records of plant phenology in China and have been instrumental in estimating change in the timing of spring plant activity in Asia (Ge et al. 2014; Vitasse et al. 2022).

The observations of *C. canadensis* submitted through *Nature's Notebook* evaluated in this study were largely contributed as a part of a formal data collection campaign coordinated by the USA National Phenology Network. The employment of the USA-NPN Redbud Phenology Project campaign was instrumental in the establishment of our flowering models, as it vastly increased the quantity of redbud flowering data and allowed us to be more selective in filtering to retain only high-quality observations. Participants were supported through multiple in-person and online training sessions and webinars and

provided with ample opportunity to clarify questions regarding flowering status; all of these efforts served to boost accuracy in the observations of flowering status.

Our work here also shows that there is potential to integrate across citizen science datasets, although how to jointly model incidental and semi-structured monitoring data in phenology is still nascent. Our work here focuses more on comparisons between herbarium and iNaturalist records and modeled results based on NPN data. These comparisons are likely to become more common and have their own inherent challenges of scale, and which phenophases are being compared. Our work is a starting point for hopefully more work focused on this integration.

Limitations and opportunities

Despite the strong performance of our hybrid modeling approach, several limitations constrain its broader interpretation and application. First, we used a pool of 179 site-year records in model construction and our first phase of validation. Such a small sample size reduced spatial and temporal replication and may have limited our ability to capture finer-scale variation, particularly in underrepresented regions such as the South and Midwest. Further, our approach, while combining mechanistic and ML-based correction layers, still excludes certain potentially important cues such as soil temperature, photothermal units, or microclimatic variation, which might further refine bloom predictions, especially in heterogeneous landscapes. That said, the successful validation of the model using historical data, which yielded predictions consistent with the expected range of flowering dates, strengthened confidence in its generalizability and in the stability of the underlying climatic relationships represented.

An important takeaway from this research was that while the focused campaign to generate observations specific to *C. canadensis* provided a robust pool of data to work with, additional observations, especially in southern states, would serve to further refine and improve predictive capabilities.

Our work opens several promising avenues for future research and application. First, the hybrid modeling approach demonstrated here—combining physiological process models with machine learning correction layers—could be readily applied to other tree species with similar ecological roles and flowering dynamics, particularly those with broad geographic ranges and volunteer-monitoring data. Second, future work could incorporate genetic or cultivar-level information to refine flowering models and explore local adaptation. Studies like Ony et al. (2021) highlight the feasibility of such integration, which would allow for climate response modeling at the population level. Third, with additional validation, this model could better support flowering forecasts for horticultural planning, pollinator support strategies, or educational

outreach. The visible and charismatic nature of redbud makes it ideal for citizen science engagement and environmental education.

Acknowledgements

We are very grateful to the many *Nature's Notebook* and iNaturalist observers who have shared their observations of flowering in *C. canadensis* and many other species. Portions of this work have been funded by the National Science Foundation, in particular grant DBI2223512 to RPG. TMC acknowledges support from the U.S. Geological Survey, Cooperative Agreement GA23AC00566. Thanks to Michelle Gelhausen, Agusti'n Baldioli, and numerous venues and media outlets for assistance promoting the Redbud Phenology Project.

Data Availability

Data files and code are available at https://github.com/tariqmohammad/C.canadensis-Flowering

References

Adkins CR, Ward NA, Braman SK, White SA (2012) Chapter 12. Redbud – *Cercis* spp., pp. 287–303. In: Flucher AF, White SA (eds) *IPM for Select Deciduous Trees in Southeastern US Nursery Production*. Southern Nursery IPM Working Group, Knoxville, Tennessee, USA, 326 pp. https://web.archive.org/web/20160212111703/http://www.clemson.edu/extension/horticulture/nursery/ipm/ipm_book.htm

Akiba T, Sano S, Yanase T, Ohta T, Koyama M (2019) Optuna: A next-generation hyperparameter optimization framework. In: *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2623–2631. https://doi.org/10.1145/3292500.3330701

Ashby W, Bresnan D, Roth P, et al. (1992) Nursery establishment, phenology, and growth of silver maple related to provenance. *Biomass and Bioenergy* 3:1–7.

Barnett PE, Farmer RE (1980) Altitudinal variation in juvenile characteristics of southern Appalachian black cherry (*Prunus serotina* Ehrh.). *Silvae Genetica* 29:157–160.

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

Belitz M, Larsen E, Reis L, Guralnick R (2020) The accuracy of phenology estimators for use with sparsely sampled presence-only observations. *Methods in Ecology and Evolution* 11(10):1273–1285.

Belitz, M. W., Larsen, E. A., Shirey, V., Li, D., & Guralnick, R. P. (2022). Phenological research based on natural history collections: Practical guidelines and a lepidopteran case study. Functional Ecology. https://doi.org/10.1111/1365-2435.14173

Brakie M (2010) Plant fact sheet for eastern redbud (*Cercis canadensis*). USDA–Natural Resources Conservation Service, East Texas Plant Materials Center, Nacogdoches, TX, USA.

Brewer S, Santiago-Blay JA (2022) The Redbud Phenology Project: Fall 2022 update for citizen scientists and anyone interested. *Life: The Excitement of Biology* 10(1):66–68. https://blaypublishers.com/wp-content/uploads/2022/11/brewer_santiago-blay_leb_101.pdf

661 Camilo GR, Fogel NS, Mullikin JC, Moss A, Edens-Meier RM, Bernhardt P (2025) Bees in a heat island:
662 bee assemblages on *Cercis canadensis* and *Cornus florida* in urban and exurban sites. *Journal of*
663 *Pollination Ecology* 38:110–131. https://doi.org/10.26786/1920-7603(2025)822

664 Caprio JM (1966) Pattern of plant development in the western United States. *Montana Agricultural*
665 *Experiment Station Bulletin* 607, Montana State University, Bozeman.

666 Chuine I (2000) A unified model for budburst of trees. *Journal of Theoretical Biology* 207(3):337–347.
667 https://doi.org/10.1006/jtbi.2000.2178

668 Chung U, Mack L, Yun JI, Kim SH (2011) Predicting the timing of cherry blossoms in Washington, DC
669 and mid-Atlantic states in response to climate change. *PLoS ONE* 6(11):e27439.
670 https://doi.org/10.1371/journal.pone.0027439

671 Crimmins TM, Marsh RL, Switzer J, Crimmins MA, Gerst KL, Rosemartin AH, Weltzin JF (2017) USA
672 National Phenology Network gridded products documentation. *US Geological Survey Open-File Report*
673 2017–1003. https://doi.org/10.3133/ofr20171003

674 Crimmins TM, Santiago-Blay JA, Brewer S, Tighe M, Posthumus EE (2024) Patterns in flowering,
675 leafing-out, and fruit ripening in eastern redbud, *Cercis canadensis* Linnaeus, 1753 (Fabaceae)
676 documented through the Nature's Notebook campaign, The Redbud Phenology Project: 2021 to the
677 present. *Life: The Excitement of Biology* 12(1):5–16.

678 Das S, Suganthan PN (2011) Differential Evolution: A survey of the state-of-the-art. IEEE Transactions
679 on Evolutionary Computation 15(1):4–31. https://doi.org/10.1109/TEVC.2010.2059031

680 Denny EG, Gerst KL, Miller-Rushing AJ, Tierney GL, Crimmins TM, Enquist CAF, Guertin P,
681 Rosemartin AH, Schwartz MD, Thomas KA, Weltzin JF (2014) Standardized phenology monitoring
682 methods to track plant and animal activity for science and resource management applications.
683 *International Journal of Biometeorology* 58:591–601. https://doi.org/10.1007/s00484-014-0789-5

684 Delpierre N, Vitasse Y, Chuine I, et al. (2016) Temperate and boreal forest tree phenology: from organ-
685 scale processes to terrestrial ecosystem models. *Annals of Forest Science* 73:5–25.
686 https://doi.org/10.1007/s13595-015-0477-6

687 Dinnage R, Grady E, Neal N, Deck J, Denny E, Walls R, Seltzer C, Guralnick R, Li D (2025)
688 PhenoVision: A framework for automating and delivering research-ready plant phenology data from field
689 images. *Methods in Ecology and Evolution*. https://doi.org/10.1111/2041-210X.70081

690 Dirr, M. A. (Illustrations by B. Dirr, M. Stephan, A. Sadauskas, and N. Snyder) (2009) Manual of Woody
691 Landscape Plants: Their Identification, Ornamental Characteristics, Culture, Propagation and Uses. 6th
692 edn. Stipes Publishing, Champaign, Illinois, USA, 1325 pp.
693
694 Dixon J, Larcenaire C, Santiago-Blay JA, Turcotte R (2024, September) Learn more about the redbud
695 seed beetle. Pest Alert. USDA Forest Service, Eastern Region, Private and Tribal Forestry, Milwaukee,
696 Wisconsin, USA. R9-PR-014-24, 2 pp.
697
698 Feldman RE, Žemaitė I, Miller-Rushing AJ (2018) How training citizen scientists affects the accuracy
699 and precision of phenological data. International Journal of Biometeorology 62:1–15.
700 https://doi.org/10.1007/s00484-018-1540-4
701
702 Fishman S, Erez A, Couvillon GA (1987) The temperature-dependence of dormancy breaking in plants:
703 Simulation of a two-step model involving a cooperative transition. Journal of Theoretical Biology
704 124(4):473–483. https://doi.org/10.1016/S0022-5193(87)80237-0

705 Friedman JH (2001) Greedy function approximation: A gradient boosting machine. Annals of Statistics
706 29(5):1189–1232. https://doi.org/10.1214/aos/1013203451

707 Fu YSH, Campioli M, Vitasse Y, De Boeck HJ, Van den Berge J, AbdElgawad H, Asard H, Piao S,
708 Deckmyn G, Janssens IA (2014) Variation in leaf flushing date influences autumnal senescence and next
709 year's flushing date in two temperate tree species. Proceedings of the National Academy of Sciences of
710 the USA 111(20):7355–7360. https://doi.org/10.1073/pnas.1321727111

711 Fu YH, Zhao H, Piao S, Peaucelle M, Peng S, Zhou G, Ciais P, Huang M, Menzel A, Peñuelas J, Song Y,
712 Vitasse Y, Zeng Z, Janssens IA (2015) Declining global warming effects on the phenology of spring leaf
713 unfolding. Nature 526(7571):104–107. https://doi.org/10.1038/nature15402

714 Fuccillo KK, Crimmins TM, DeRivera C, Elder TS (2015) Assessing accuracy in citizen science-based
715 plant phenology monitoring. International Journal of Biometeorology 59:917–926.
716 https://doi.org/10.1007/s00484-014-0892-7

Gallinat AS, Schwartz MD, Donnelly A, Li X, Crimmins TM (2025) Combined volunteer and ecological network observations show broad-scale temperature-sensitivity patterns for deciduous plant flowering and leaf-out times across the Eastern USA. Journal of Ecology. (In press)

Ge Q, Wang H, Zheng J, This R, Dai J (2014) A 170-year spring phenology index of plants in eastern China. Journal of Geophysical Research: Biogeosciences 119:301–311. https://doi.org/10.1002/2013JG002565

Gleason HA, Cronquist A (1991) Manual of Vascular Plants of Northeastern United States and Adjacent Canada. 2nd edition. Bronx, NY, USA.

Hamel PB, Chiltoskey MU (1975) Cherokee Plants and Their Uses: A 400 Year History. Herald Publishing Co., Sylva, North Carolina, USA, 65 pp. https://archive.org/details/cherokeeplantsth0000paul

Hamel PB, Chiltoskey MU (2002) Cherokee Plants and Their Uses: A 400 Year History. Book Publishing Company, Sylva, North Carolina, USA, 71 pp. https://archive.org/details/cherokeeplantsth0000paul

Hänninen H (1990) Modelling bud dormancy release in trees from cool and temperate regions. Acta Forestalia Fennica 213:7660. https://doi.org/10.14214/aff.7660

Hatzis JJ, Schwartz MD, Ault TR, Donnelly A, Gallinat A, Li X, Crimmins TM (2025) Building Spring Development Indices for woody species in the conterminous United States. Agricultural and Forest Meteorology 110443. https://doi.org/10.1016/j.agrformet.2025.110443

Hawkins CDB, Dhar A (2012) Spring bud phenology of 18 Betula papyrifera populations in British Columbia. Scandinavian Journal of Forest Research 27:507–519.

Herms DA (2004) Using degree-days and plant phenology to predict pest activity. Tactics and Tools for IPM 58:49–59.

Hopkins AD (1920) The bioclimatic law. Journal of the Washington Academy of Sciences 10(2):34–40.

Hopp RJ (1974) Plant phenology observation networks. In: Phenology and Seasonality Modeling, 25–43. Springer, Berlin Heidelberg.

Hough FB (1864) Observations upon periodical phenomena in plants and animals from 1851 to 1859. Executive Documents, 36th Congress, 1st Session, Washington, DC.

Jefferson T (1791, May 8) Letter to Mary Jefferson. Founders Online, National Archives. https://founders.archives.gov/documents/Jefferson/01-20-02-0121

Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, Ye Q, Liu T-Y (2017) LightGBM: A highly efficient gradient boosting decision tree. In: Advances in Neural Information Processing Systems 30:3146–3154.

Koeser AK, Hassing G, Friedman MH, Irving RB (2022) Eastern redbud. In: Trees: North & Central Florida. A Field Guide to 140 Tree Species. Institute of Food and Agricultural Sciences, University of Florida, Gainesville, FL, USA.

Kimura K, Yasutake D, Oki T, Yoshida K, Kitano M (2021) Dynamic modelling of cold-hardiness in tea buds by imitating past temperature memory. Annals of Botany 127(3):317–326. https://doi.org/10.1093/aob/mcaa197

Larsen EA, Belitz MW, Guralnick RP, Ries L (2022) Consistent trait–temperature interactions drive butterfly phenology in both incidental and survey data. Scientific Reports 12(1):13370.

Laube J, Sparks TH, Estrella N, Höfler J, Ankerst DP, Menzel A (2014) Chilling outweighs photoperiod in preventing precocious spring development. Global Change Biology 20(1):170–182. https://doi.org/10.1111/gcb.12360

Leys C, Ley C, Klein O, Bernard P, Licata L (2013) Detecting outliers: do not use standard deviation around the mean, use absolute deviation around the median. Journal of Experimental Social Psychology 49(4):764–766. https://doi.org/10.1016/j.jesp.2013.03.013

Liang L (2016) Beyond the bioclimatic law: geographic adaptation patterns of temperate plant phenology. Progress in Physical Geography 40(6):811–834.

Liang L (2025) Mapping range-wide budburst gradients in 22 temperate tree species across the eastern U.S. using quantile regression. Agricultural and Forest Meteorology 373:110790. https://doi.org/10.1016/j.agrformet.2025.110790

Lovell HB (1961) Let's talk about honey plants. Gleanings in Bee Culture 89(2):99–100.

776 Luedeling E, Brown PH (2011) A global analysis of the comparability of winter chill models for fruit
777 and nut trees. International Journal of Biometeorology 55(3):411–421. https://doi.org/10.1007/s00484-
778 010-0352-y
779
780 Luedeling E (2012) Climate change impacts on winter chill for temperate fruit and nut production: A
781 review. Scientia Horticulturae 144(9):218–229. https://doi.org/10.1016/j.scienta.2012.07.011
782
783 Lundberg SM, Lee S-I (2017) A unified approach to interpreting model predictions. In: Advances in
784 Neural Information Processing Systems 30:4765–4774. (Preprint: https://arxiv.org/abs/1705.07874)

785 McCormick RF, et al. (2021) Intercontinental prediction of soybean phenology via a hybrid ensemble of
786 knowledge-based and data-driven models. In Silico Plants 3(1):diab004.
787 https://doi.org/10.1093/insilicoplants/diab004

788 Morin X, Lechowicz MJ, Chuine I (2009) Leaf phenology in 22 North American tree species: Do spatial
789 patterns reflect climatic adaptation? Global Change Biology 15(4):961–975.
790 https://doi.org/10.1111/j.1365-2486.2008.01735.x

791 Nelder JA, Mead R (1965) A simplex method for function minimization. The Computer Journal
792 7(4):308–313. https://doi.org/10.1093/comjnl/7.4.308

793 Ony MA, Klingeman WE, Zobel J, Trigiano RN, Ginzel M, et al. (2021) Genetic diversity in North
794 American Cercis canadensis reveals an ancient population bottleneck that originated after the last glacial
795 maximum. Scientific Reports 11(1):21803. https://doi.org/10.1038/s41598-021-01020-z

796 Pearse WD, Davis CC, Inouye DW, Primack RB, Davies J (2017) A statistical estimator for determining
797 the limits of contemporary and historic phenology. Nature Ecology & Evolution 1:1876–1882.
798 https://doi.org/10.1038/s41559-017-0350-0

799 Perrier A, Turner MC, Galloway LF (2025) Shifts in vernalization and phenology at the rear edge hold
800 insight into the adaptation of temperate plants to future milder winters. New Phytologist 246:1377–1389.
801 https://doi.org/10.1111/nph.70005

802 Posthumus E, Crimmins T, Santiago-Blay JA (2021) Participate in citizen science: The Redbud Project.
803 Wednesday, January 5, 2022, 12 Noon Eastern USA Time. Life: The Excitement of Biology 9(1):27–28.
804 https://blaypublishers.com/wp-content/uploads/2021/12/posthumus_leb9131dec.21.pdf

PRISM Climate Group (2024) PRISM gridded climate data, version 2.1. Oregon State University. http://prism.oregonstate.edu

Reader R, Radford JS, Lieth H (1974) Modeling important phytophenological events in Eastern North America. In: Phenology and Seasonality Modeling. Springer, Berlin Heidelberg.

Richardson AD, Hufkens K, Milliman T, et al. (2018) Ecosystem warming extends vegetation activity but heightens vulnerability to cold temperatures. Nature 560(7718):368–371. https://doi.org/10.1038/s41586-018-0399-1

Richardson EA, Seeley SD, Walker DR (1974) A model for estimating the completion of rest for 'Redhaven' and 'Elberta' peach trees. HortScience 9(4):331–332.

Robertson KR (1976) *Cercis*: The redbuds. Arnolodia 36(2):37-49. https://doi.org/10.5962/p.249677

Rodríguez A, Álvarez I, Spano D, Cantão JP (2019) Chilling accumulation in fruit trees in Spain under climate change. Natural Hazards and Earth System Sciences 19:1087–1103.

Storn R, Price K (1997) Differential Evolution—A simple and efficient heuristic for global optimization over continuous spaces. Journal of Global Optimization 11:341–359. https://doi.org/10.1023/A:1008202821328

Taylor SD, Meiners JM, Riemer K, Orr MC, White EP (2019) Comparison of large-scale citizen science data and long-term study data for phenology modeling. Ecology 100(2):e02568. https://doi.org/10.1002/ecy.2568

Teuton C, Shade H, with Shade L, Shade L; illustrated by M. Timothy (2023) Cherokee Earth Dwellers: Stories and Teachings of the Natural World. University of Washington Press, Seattle, WA, USA, 296 pp.

Tighe M, Brewer S, Santiago-Blay JA (2024) The redbud, Cercis canadensis Linnaeus, 1753 and Cercis occidentalis (Fabaceae) Torrey ex A. Gray (1850), Phenology Project. Life: The Excitement of Biology 11(3):96–99. https://blaypublishers.com/wp-content/uploads/2024/02/tighe_et_al_leb_113-2.pdf

USA National Phenology Network (2025) Plant and animal phenology data (Individual Phenometrics), 2011–2025 for region 28.8°N to 49.9°N, contiguous United States. USA-NPN, Tucson, Arizona, USA. Accessed 20 July 2025. https://doi.org/10.5066/F78S4N1V

van Bree R, et al. (2025) Hybrid phenology modeling for predicting temperature effects on tree dormancy. arXiv preprint arXiv:2501.16848. https://arxiv.org/abs/2501.16848

Vitasse Y, Porté AJ, Kremer A, Michalet R, Delzon S (2009) Responses of canopy duration to temperature changes in four temperate tree species: relative contributions of spring and autumn leaf phenology. Oecologia 161(1):187–198. https://doi.org/10.1007/s00442-009-1363-4

Vitasse Y, Baumgarten F, Zohner CM, Rutishauser T, Pietragalla B, Gehrig R, Dai J, Wang H, Aono Y, Sparks TH (2022) The great acceleration of plant phenological shifts. Nature Climate Change 12(4):300-302. https://doi.org/10.1038/s41558-022-01283-y

Way DA, Montgomery RA (2015) Photoperiod constraints on tree phenology, performance, and migration in a warming world. Plant, Cell & Environment 38(9):1725–1736. https://doi.org/10.1111/pce.12431

Wang H, Wu C, Ciais P, Penuelas J, Dai J, Fu Y, Ge Q (2020a) Overestimation of the effect of climatic warming on spring phenology in temperate forests. Nature Communications 11:4945. https://doi.org/10.1038/s41467-020-18743-8

Wang H, Wang H, Ge Q, Dai J (2020) The interactive effects of chilling, photoperiod, and forcing temperature on flowering phenology of temperate woody plants. Frontiers in Plant Science 11:443. https://doi.org/10.3389/fpls.2020.00443

Werle CT, Witcher AL (2002) Potential impacts of shade treatments on dormancy of overwintering redbud (*Cercis canadensis* L.) trees at southeastern nurseries. *Journal of Environmental Horticulture* 40(2):79–86. https://doi.org/10.24266/2573-5586-40.2.79

Wood RJ, Orel V (2001) Genetic Prehistory in Selective Breeding: A Prelude to Mendel. Oxford University Press, Oxford, UK, 323 pp.

Zhang Y, Steiner AL (2022) Projected climate-driven changes in pollen emission season length and magnitude over the continental United States. Nature Communications 13(1):1234.

Zirkle C (1935) *The Beginnings of Plant Hybridization*. Morris Arboretum Monographs I. University of Pennsylvania Press, Philadelphia, PA, USA, 231 pp.

Zohner CM, Renner SS (2017) Innately shorter vegetation periods in North American species explain native–non-native phenological asymmetries. *Nature Ecology & Evolution* 1:1655–1660. https://doi.org/10.1038/s41559-017-0307-3