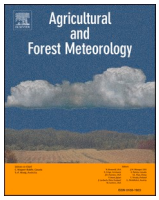
Agricultural and Forest Meteorology 364 (2025) 110443

Contents lists available at ScienceDirect

Agricultural and Forest Meteorology

journal homepage: www.elsevier.com/locate/agrformet



Building spring development indices for woody species in the conterminous United States

Joshua J. Hatzis [a], Mark D. Schwartz [a,*], Toby R. Ault [b], Alison Donnelly [a], Amanda Gallinat [c], Xiaolu Li [b], Theresa M. Crimmins [d,e]

[a] Department of Geography, University of Wisconsin-Milwaukee, Milwaukee, WI, USA
[b] Department of Earth and Atmospheric Sciences, Cornell University, Ithaca, NY, USA
[c] Department of Environmental Studies, Colby College, Waterville, ME, USA
[d] School of Natural Resources and the Environment, University of Arizona, Tucson, AZ, USA
[e] USA National Phenology Network, Tucson, AZ, USA

ARTICLE INFO

Keywords:
Phenology
Phenoclimatology
Modeling
Indices
Conterminous United States
Gridded climate data

ABSTRACT

Phenological indices are an effective approach for assessing spatial and temporal patterns and variability in plant development. The Spring Indices (SI-x), two widely adopted phenological indices, have been used in recent decades to predict development of woody plants, and document changes in spring growth timing, especially in North America. However, these two indices (Leaf and Bloom) capture only two "moments" in the continuum of spring when quantities of thermal or photo/thermal energy, associated with seasonal events in plants, are accumulated, limiting their utility to characterize the remainder of the spring season. Further, the Spring Indices do not account for intraspecific variation, limiting their ability to reflect non-cloned plant development. To address these shortcomings, we developed a novel suite of phenological indices that encompass a broader span of the spring season. These indices were constructed using observations contributed to the USA National Phenology Network's *Nature's Notebook* platform across many non-cloned tree and shrub species' ranges, thereby incorporating differing regional responses within species due to genetic variations.

Individual species model predictions of leaf or bloom timing exhibited an average mean absolute error of 8.55 days; most were improved by the inclusion of site-specific latitude, elevation, or 30-year average temperature. Leaf and bloom model outputs for individual species across the spring season were temporally aggregated into four leaf and bloom groups to produce a suite of Spring Development Indices (SDI). Accuracy of the SDI predictions was 0.89 days lower, on average, than the species models, but 2.65 days better than SI-x. Generally, all SDIs were highly correlated. The SDIs exhibiting the most difference from the others were Early leaf, Very Early bloom, and Late bloom. As such, these SDIs provide novel insights, beyond SI-x, into the relative timing of spring-season "moments" across species in space and time.

1. Introduction

Phenology, the timing of seasonal developmental phases of plants and animals as driven by environmental factors, plays a major role in governing nutrient cycling, species interactions, and ecosystem services. The timing of seasonal events in plants and animals also affects organismal growth, performance, and fitness (Chuine and Régnière 2017). Accordingly, phenological models, which predict the timing of seasonal events such as leaf and bloom in plants, and egg hatch and adult emergence in insects, have become increasingly useful in applications such as scheduling pest and invasive species treatment, predicting development in agricultural species, and in modeling population dynamics (e.g., Adole et al., 2019; Crimmins et al. 2020; Park et al. 2019; Roslin et al. 2021; Taylor and White 2020).

A strength of phenological models is that they can be used to predict the timing of seasonal activity or a seasonal transition in a selected organism at any location and any time (present, past, or future) for which environmental variable data are available. As such, phenological models can also be used to evaluate changes in phenology—or when particular conditions associated with particular biological events are

* Corresponding author at: Geography Department, UW-Milwaukee, P.O. Box 413, Milwaukee, WI 53201, USA.
E-mail address: mds@uwm.edu (M.D. Schwartz).

https://doi.org/10.1016/j.agrformet.2025.110443
Received 29 May 2024; Received in revised form 27 January 2025; Accepted 9 February 2025
0168-1923/© 2025 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

reached—driven by changing climate conditions (e.g., Blümel and Chmielewski 2012; Chuine et al. 1999; Črepinšek et al. 2006; Hänninen 1990; Masle et al. 1989).

Most phenological models predict the timing of a single seasonal event, or phenophase, in a single species. However, a handful of phenological indices exist, which strive to predict the timing of a community- or ecosystem-level phenomenon (e.g., Jolly et al. 2005). The Spring Indices are a widely implemented set of phenology indices in the United States. The original Spring Indices (SI-o) were developed using observations of leafing and blooming in cloned lilac and honeysuckle plants, and incorporated chilling requirements (Ault et al. 2015a; Schwartz et al. 2006, 2012). Schwartz et al. (2013a) removed the chilling requirement, demonstrating that the "extended" models (SI-x) were equally valid, and able to be calculated across the entire conterminous USA. The Spring Indices have been used to reveal past variations in the timing of spring growth (Schwartz 1994, 1998; Schwartz and Reiter 2000; Schwartz et al. 2006), document changes in spring phenology for horticulturally relevant woody perennials in the northeastern U.S. (Wolfe et al. 2005), suggest potential advancements of spring onset timing for temperature-sensitive plants under future climate change scenarios (Li et al. 2023), and explore "false spring" frequency in the 21st century under projected climate change (Allstadt et al. 2015).

Two distinct indices comprise the Spring Indices: a Leaf Index and a Bloom Index. The Leaf Index reflects the average date leaves first emerge on the three plant species (two honeysuckles and a lilac), and the Bloom Index indicates the average date blooming begins in these plants. The output of each index is the day of year that a particular set of environmental conditions are reached. As such, each index can be viewed as reflecting a particular "moment" in the continuous progression of the spring season, related to when a quantity of thermal or photo/thermal energy accumulates to a fixed threshold. An even simpler example of a "moment" in spring is a particular number of growing degree-hours accumulated following a fixed start date.

The "moments" in the progression of the spring season captured by the Spring Leaf and Bloom Indices can be used to infer when other early-season species undergo seasonal events. The Spring Indices have proven capable of predicting leaf emergence and bloom with reasonable accuracy for temperature-responsive early season trees and shrubs (Gerst et al. 2020; Schwartz et al. 2006; Schwartz et al. 2013a). However, these indices are based on only three genetically identical, or cloned, plant species. Natural (i.e., not cloned) species often exhibit varying responses to environmental variables across their ranges due to genetic variation, something that SI-x does not address (Liang and Schwartz 2014; Liang 2019). Further, SI-x represents only two discrete moments that occur relatively early in the continuous progression of the spring season. And while these indices can be useful for indicating whether these early moments in spring are being reached earlier or later than normal in a given year, they offer only limited information regarding how the rest of the season may play out for other species undergoing leaf-out and bloom later in the season. As Crimmins and Crimmins (2019) have demonstrated, an early start to the season does not necessarily indicate a continuation of early activity throughout the season. To more fully characterize plant activity over the course of the spring season, and whether conditions associated with activity in plants are being reached earlier or later than average in a particular year, additional indices that use many species to encompass more of the spring season are needed.

In this study, we develop a new suite of phenological indices distributed temporally across the entirety of the spring season, using observations of natural (not cloned) plants. We follow the general approach undertaken in the development of SI-x, of first developing models predicting leaf budbreak and open flowers for individual species and then aggregating these models into synthetic indices. Indices represent when quantities of thermal or photo/thermal energy are accumulated to fixed thresholds that reflect conditions associated with a group of plants undergoing seasonal events at approximately the same time. Our approach to developing indices is comparable to an established technique referred to as a "phenological calendar" (Henniges et al. 2005; Nekovar et al. 2008). In that technique a small number of events like "breaking leaf buds" and "open flowers" are tracked for multiple species from the earliest to the latest part of the spring season. Rather than using additional phenological events such as fruit development and ripening within these same species to constitute the "progression" of spring, this approach uses the differing timing of the same events for a larger pool of species. Thus, phenological calendars can establish the typical timing of spring plant development across the spring season calendar (and be sensitive to any changes), by employing observations of multiple species. Therefore, our new indices will encompass more of the span of time that comprises the spring calendar season. As such, these new indices will expand our ability to track annual spring development across a wider variety of species and ecosystems in North America and contribute to better understanding of the variable influences of climate change on spring plant growth.

2. Data and methodology

2.1. Data

In this study, we utilize observations of leaf and bloom activity contributed across the United States to the USA National Phenology Network's *Nature's Notebook* platform. These phenological observations are collected by professional and volunteer observers (Rosemartin et al. 2018) following robust, scientifically vetted observation protocols (Denny and Crimmins 2023; Denny et al. 2014). Presently, 1458 species of plants are available for monitoring through *Nature's Notebook*; approximately 40 % of these are deciduous trees and shrubs (USA-NPN 2024).

An overview of the data screening steps, and model construction process are diagramed in Fig. 1. Status observations of "breaking leaf buds" (leaf, USA-NPN Phenophase 371, referred to going forward as "leaf") and "open flowers" (bloom, USA-NPN Phenophase 501, referred to going forward as "bloom"), over 2010–2021 were acquired from the USA National Phenology Network (USA-NPN, Schwartz et al. 2012; USA-NPN 2022). Notably, phenological data collected by the USA National Ecological Observation Network (NEON) sites are incorporated into the USA-NPN database (Elmendorf et al. 2016). We took three steps to improve the quality of the data used. First, all records were removed where a report of "yes" to "breaking leaf buds" or "open flowers" was not preceded by a "no" report in the preceding 14 days. Second, we removed observations with conflicting phenological status reports for the same individual plant on the same day and sites exhibiting >5 % of observations in conflict. Lastly, we removed observations of leaf or bloom that occurred after day-of-year 200 (mid-summer, ~July 19) and those which belonged to non-woody species. We next took four steps (see Fig. 1) to improve the likelihood that each leaf or bloom species data set would be large and spatially distributed to support development of a model to be applied continent-wide. In this process we removed species with any of the following: 1) fewer than 100 total observations; 2) observations at fewer than 10 sites; 3) observations spread over a total area of <100,000 km^2; or 4) observations confined to a single climate region (Fig. 2a). After these filters, we retained 46 species with leaf data and 41 species with bloom data (Fig 2a). Since leaf and bloom models for species were developed independently it was not necessary to have both leaf and bloom data for individual species.

For each observation location and date, we downloaded 4 km Parameter-elevation Regressions on Independent Slopes Model (PRISM) gridded daily weather datasets (Daly et al. 2008; Daly et al. 2021; PRISM Climate Group 2023; Stoklosa et al. 2015) over 1991–2022 for the following variables: mean, minimum, and maximum daily air temperature, minimum and maximum vapor pressure deficit, daily precipitation, and elevation. We used this longer time period (32 years compared to only 12 years of phenological data) to allow projection of our model

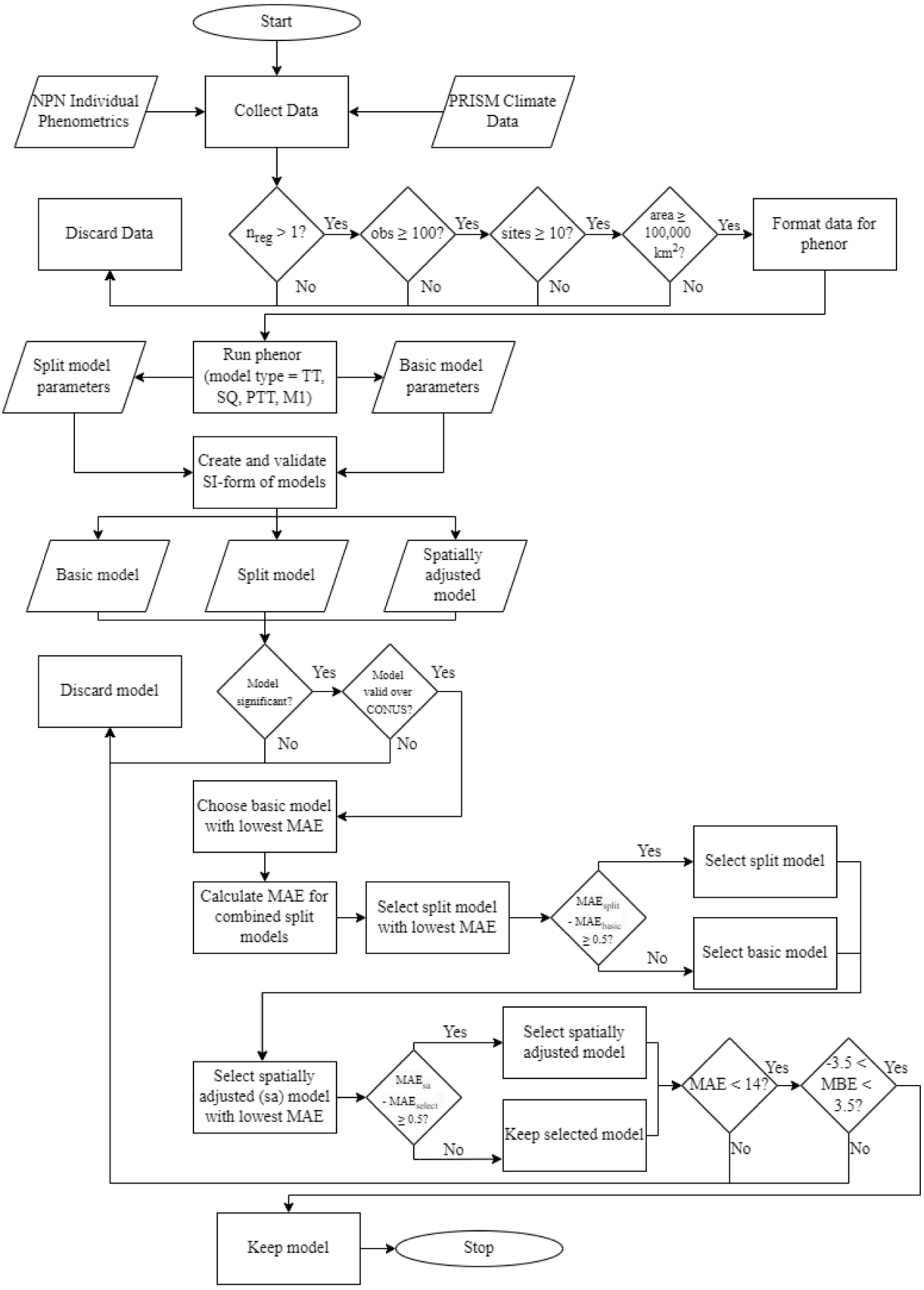


Fig. 1. Flow diagram of the species selection and construction processes for each final leaf or bloom species model (henceforth termed Species x phenophase models).

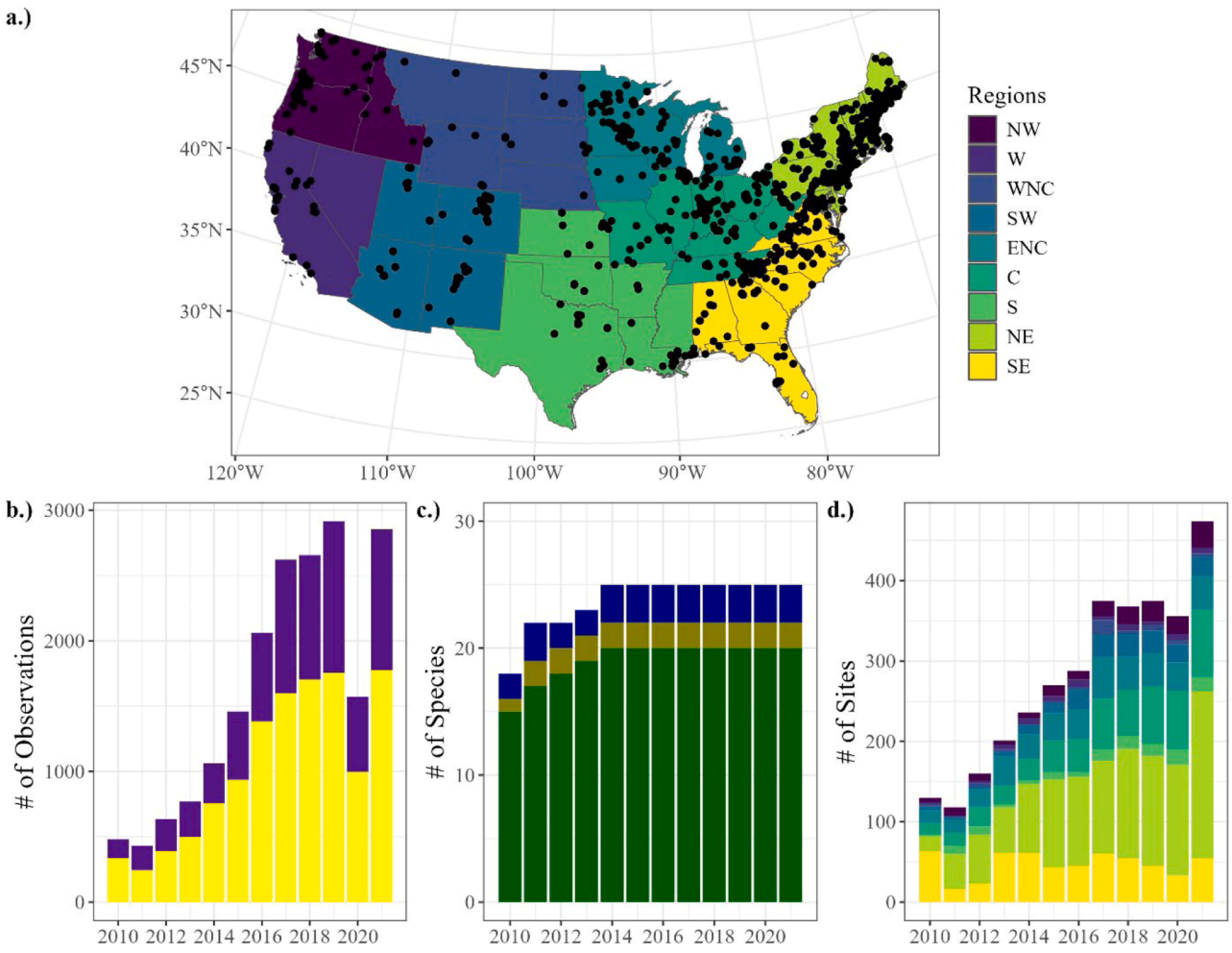


Fig. 2. Location of USA-NPN observation sites and the National Centers for Environmental Information (NCEI) Climate Regions of the conterminous USA, where the regions are Northwest (NW), West (W), West North Central (WNC), Southwest (SW), East North Central (ENC), Central (C), South (S), Northeast (NE), and Southeast (SE) (panel 2a). Number of leaf (yellow) and bloom (purple) observations available per year (panel 2b). Number of species available per year per lifeform type (Govaerts et al. 2021; Plants of the World Online 2023): tree (green), tree or shrub (dark yellow), shrub (dark blue) (panel 2c). Number of observation sites per year per climate region (panel 2d), with colors as in panel 2a.

results over the entire current standard 30-year climatological time period, as well as to the most recent PRISM data available. Daylength (photoperiod) was calculated for each day and location as a function of latitude and day of year (DOY), following Forsythe et al.'s (1995) method. For purposes of validation and comparison, we also downloaded the following gridded data sets (4km) from the USA-NPN Geoserver: extended Spring Indices – First Bloom (Using PRISM Data; USA-NPN 2023a) and extended Spring Indices – First Leaf (Using PRISM Data; USA-NPN 2023b) for the period Jan. 1, 1991, to Dec. 31, 2022.

2.2. Methodology

2.2.1. Species × phenophase model selection

We first randomly split our leaf and bloom data into training (60 %) and validation (40 %) datasets, which is a standard practice for model development (Joseph 2022). We then fitted the data to one of the four simplest model types available in the *phenor* package (Hufkens et al. 2018) for R (R Core Team 2022): thermal time (TT), sequential (SQ), photothermal time (PTT), and M1 photothermal time (M1). These model types are well understood and tested, and all their parameters can be generated from daily maximum-minimum temperature and daylength data. The thermal time model calculated each phenophase occurrence as the day-of-year when the accumulated growing degree hours (Ault et al. 2015b) reached a function-specific threshold (Réaumur 1735; Wang 1960). The sequential model had a chilling requirement (accumulated chill units) that must be met before growing degree hours could begin to accumulate. Once growing degree hours began accumulating the phenophase date was considered the day-of-year when the function-specific threshold was met (Hänninen 1990; Kramer 1994). The photothermal time model resembled the thermal time model except growing degree hours were moderated by the fraction of daylight hours (Črepinšek et al. 2006; Masle et al. 1989). The M1 photothermal time model was like the photothermal time model except the daylight parameter was one-tenth of the daylight hours raised to a function-specific parameter (Blümel and Chmielewski 2012). We chose to use *phenor* as it incorporates a generalized annealing process to optimize model type and parameter selection.

We next determined the best model for each species × phenophase (one with lowest mean absolute prediction error) and derived the corresponding optimal parameters for that model using *phenor*. The parameters optimized were the day when heat accumulation began (t_0; days since Oct. 1), base temperature for growing degree hour calculation (T_{base} in °C), day when chill unit accumulation began ($t_{0,\ chill}$; days since Oct. 1), minimum, optimum, and maximum temperatures for plant growth (T_{min}, T_{opt}, and T_{max} respectively, all in °C), required chill units to break dormancy (C_{req} in chill units), and power for daylight parameter in the M1 photothermal model (f). It is worth noting that in our early experiments with *phenor* we discovered that if the start date of

thermal accumulation was allowed to vary, the optimization process would often place it after the occurrence of some early phenophases. This reduced overall prediction error but excluded those events from consideration in model construction. To avoid this issue, we decided to use a fixed thermal accumulation start date of December 1st ($t_0 = 62$) for all models which was early enough to ensure no phenophase data were excluded across the conterminous USA (CONUS) for any species. Next, using the selected model type equation and optimized parameters for each species × phenophase we calculated the accumulated thermal forcing as defined by Basler (2016), on the day of the phenophase, for every location and year (Fig. 1).

The Spring Indices (SI-x) form for each species × phenophase equation was then derived (Ault et al. 2015a,b; Schwartz 1990; Schwartz et al. 2006; Schwartz et al. 2013a) to make predictions and evaluate model errors. This involved using accumulated thermal forcing (S_{frc}) as the independent variable, and phenophase (T) days since Oct. 1 as the dependent variable in a simple linear regression:

$$\frac{1000}{T} = \alpha + \beta\frac{S_{frc}}{T}, \tag{1}$$

Where 1000 is an arbitrary constant (Ault et al. 2015b) and α and β are regression coefficients. Once the regression coefficients were calculated, the equation was rearranged to estimate daily thermal forcing ($\widehat{C}$) as:

$$\widehat{C}(t) = \alpha t + \beta\frac{S_{frc}}{T}t, \tag{2}$$

Where t is the current number of days since Oct. 1. The estimated number of days since Oct. 1 for the phenophase (S_p; model prediction) was then defined as the number of days since Oct. 1 when the cumulative thermal forcing exceeded 1000:

$$S_p = \begin{cases} t, & \sum_{t=0}\widehat{C}(t) \geq 1000 \\ undefined, & otherwise \end{cases} \tag{3}$$

Finally, we converted the days from days since Oct. 1 (used in *phenor*) to days since Jan. 1 (DOY) to get the calendar dates for the phenophases. We used the SI-x framework because it offered several useful features. First, unlike conventional regression, it allowed day-of-year dates to be used directly as dependent variables at locations across large areas with different phenological dates, but the same expected independent variable accumulations (in this case 1000). Second, it provided more precision when calculating model error (mean absolute error, MAE and mean bias error, MBE) over the conventional regression approach (such as used in *phenor*). Conventional regression evaluates error by using accumulated independent variable values all at the time of each individual phenophase date (dependent variable) to produce predictions. In contrast, the SI-x framework (moving forward from the start date, see Eq. (2) above) more realistically incorporates new independent variable value accumulations each day, until the model prediction time is reached. Thus, predictions are determined without directly involving phenophase dates, and model errors calculated using independent variable (accumulated thermal forcing) values on the actual day of each prediction, rather than only at all the phenophase times (Ault et al. 2015b; Schwartz 1997; Schwartz et al. 2013a). The models produced to this point are called "basic", as they use the same model form, parameters, and regression coefficients for all locations across the entire CONUS (Fig. 1).

We were interested in testing for variations in the most accurate species model forms and parameters that could be developed for different regions across the CONUS. To evaluate this influence of geography on model accuracy we created two additional SI-form model types for each species × phenophase: "split" models and "spatially adjusted" models. The split model procedure divided the CONUS into multiple regions, but then used the same model form, parameters, and regression coefficients for all locations within each region. The process involved prior to estimating parameters in *phenor*, spliting the CONUS-wide datasets into regions based on the following observation site specific variables: 30-year mean (1991–2020) annual site temperature (MAT), and site latitude (LAT). We use both two (MAT2; $MAT < 13^{\circ}C$ and $MAT \geq 13^{\circ}C$) and three (MAT3; $MAT < 10^{\circ}C$, $10^{\circ}C \leq MAT \leq 16^{\circ}C$, $MAT > 16^{\circ}C$) groups for MAT and two (LAT2; $LAT < 40^{\circ}\ N$ and $LAT \geq 40^{\circ}\ N$) and four (LAT4; $LAT \leq 32^{\circ}\ N$, $32^{\circ}\ N < LAT \leq 37^{\circ}\ N$, $37^{\circ}\ N < LAT \leq 42^{\circ}\ N$, $LAT > 42^{\circ}\ N$) groups for LAT. We then derived the optimal parameter types for the regression models from *phenor*, as before, for each region and then created the corresponding SI-x model forms (Fig. 1).

For the spatially adjusted models, we used 30-year mean (1991–2020) annual site temperature (MAT), site latitude (LAT), and site elevation (ELEV). We first calculated bias (B) and then related it to one of the site variables (X) through a second regression:

$$B = \gamma + \delta X, \tag{4}$$

where γ and δ are regression coefficients. Once the regression coefficients were calculated, we calculated the final phenophase date as follows:

$$\widehat{S_p} = S_p - \widehat{B} \tag{5}$$

where $\widehat{B}$ is the estimated bias after correction. The spatially adjusted approach allowed the basic (CONUS-wide) and split (regional groups) models output to be improved by spatial corrections to all the point locations. This yielded 20 SI-form models for each *phenor* model type, phenophase, and species.

To test the capability of each model to produce CONUS-wide output we first resampled the long-term annual average (1991–2022) PRISM raster data sets through bilinear interpolation to a coarser resolution of 15 km. We then used each model to estimate the long-term annual average phenophase date for each grid point as well as the mean phenophase date for the entire CONUS and discarded all models that had any grid points with undefined phenophase dates. We also discarded any model that had any parameters that were not significant ($\alpha > 0.05$). Model selection was based on lowest mean absolute error (MAE), which is well regarded for its ability to comparatively assess model performance (Willmott and Matsuura 2005; Willmott et al., 2017).

From this point, we began the selection process with only the basic and split model versions for each species x phenophase (Fig. 1). We then grouped the uncorrected SI-form models by species, phenological event, and model group and selected the model with the lowest MAE.

Next, we estimated the phenophase date for each validation dataset for each model (averaging together results for the split model types across all regions) and determined MAE for each model's entire validation dataset. Fig. 3 shows the latitude zones across the conterminous USA (Fig. 3a) as a visual example of the accuracy of a model split by latitude for red maple (*Acer rubrum*; Figs. 3b-c). Finally, we selected the model with the lowest MAE from among the basic model, MAT split models, and LAT split models for each species × phenophase. To help ensure that adding additional complexity to the models improved them, we required that the split models show reduced MAE, over the basic model, by at least a half day (Fig. 1). If the split models performed worse or reduced the MAE by less than this, we selected the basic model, otherwise we selected the best split model. Next, we considered the spatially adjusted models and compared their MAEs to the best model chosen in the previous step. As before, if the spatial adjustment process improved the MAE by at least half a day we selected the best spatially adjusted model, otherwise we stuck with the previously selected base or split model (Fig. 1). As a last step, we felt it necessary to place an upper limit on the MAE and MBE thresholds, so as to not retain models with excessive errors. Given the scarcity of comparable continental-wide modeling studies, there is no consensus on the level of acceptable model error at this scale. So based on the MAE and MBE distributions of

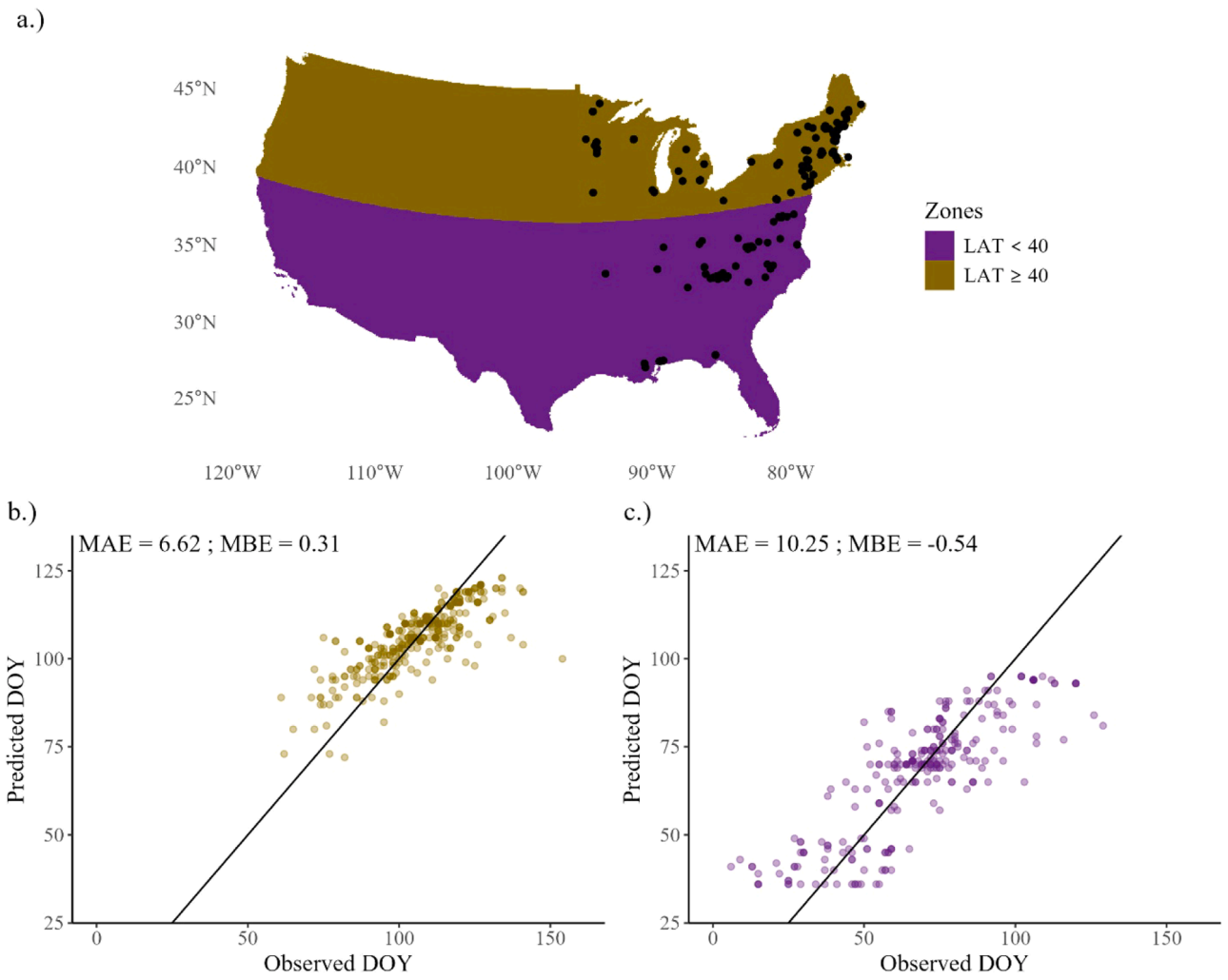


Fig. 3. Panel (a) shows the latitude zones and station locations within the conterminous USA. Predicted versus observed bloom dates are shown for the high latitude zone (LAT ≥ 40° N, panel b), and low latitude zone (LAT < 40° N, panel c), for Acer rubrum bloom models.

our selected species x phenophase model, and previous modeling experience in the development of SI, we arbitrarily removed any models with either an MAE greater than 14 days or mean bias error (MBE) less than −3.5 days or greater than 3.5 days. This yielded 24 leaf models and 27 final bloom models for use in the development of indices.

2.2.2. *Indices construction*

The process used to build our new indices is shown in Fig. 4. We used partitioning about medoids (PAM) clustering (Kaufman and Rousseeuw 1990), to create temporal clusters for the candidate leaf and bloom models: very early, early, middle, and late, and the "elbow" or within-sums-of-squares (WSS) method (Thorndike 1953) to choose the number of clusters (4). This approach allows the clusters to be evenly spread across the temporal range of species x phenophase model outputs, and provide maximum spring coverage. As a final cluster selection criterion, we first performed Pearson's correlation among all the mean annual phenophase x species dates over the entire CONUS between 1991 and 2022 within clusters and each average cluster date. In order to reduce internal dissimilarities, we removed any phenophase x species models from each cluster that had a correlation to the cluster average value of <0.9. This resulted in a final set of leaf models for 21 species and bloom models for 20 species. As with SI-x, we defined our new indices, named Spring Development Indices (SDIs), as the multi-species mean model date for each cluster of models (Schwartz 1997). For example, if one index was comprised of five species models then the index date would be the average of the five individual model dates.

2.2.3. *Spring development indices performance, separability tests, and trend analyses*

Given the lack of comparable continental scale models for comparison, we opted to evaluate SDI performance by comparing them to both their component species × phenophase model predictions and as spatially corresponding values of the two SI-x indices. We first reused the averaged 15 km yearly species × phenophase model dates for the entire CONUS over 1991–2022. We then collected the modeled dates for every species belonging to the same cluster at the same location and year and calculated the SDI date as the mean of all index member model dates for that location and year. Next, we calculated the SI-x indices DOY value for the same phenophase, location, and time. We then compared the species x phenophase model, SI-x, and SDI values (as predictions), to the available actual species x phenophase values (all species did not have data at all locations) and calculated initial MAE and MBE values.

While the species x phenophase models were calibrated directly by the multiple regression, and therefore average values were estimated based on species-level data, average values for the SI-x and SDI indices were based on multi-species mean dates for model clusters; therefore, temporal biases needed to be corrected prior to direct comparison of these outputs. This correction was accomplished by subtracting the calculated original overall average bias error from the original individual SI-x and SDI predictions at each location and year over the entire

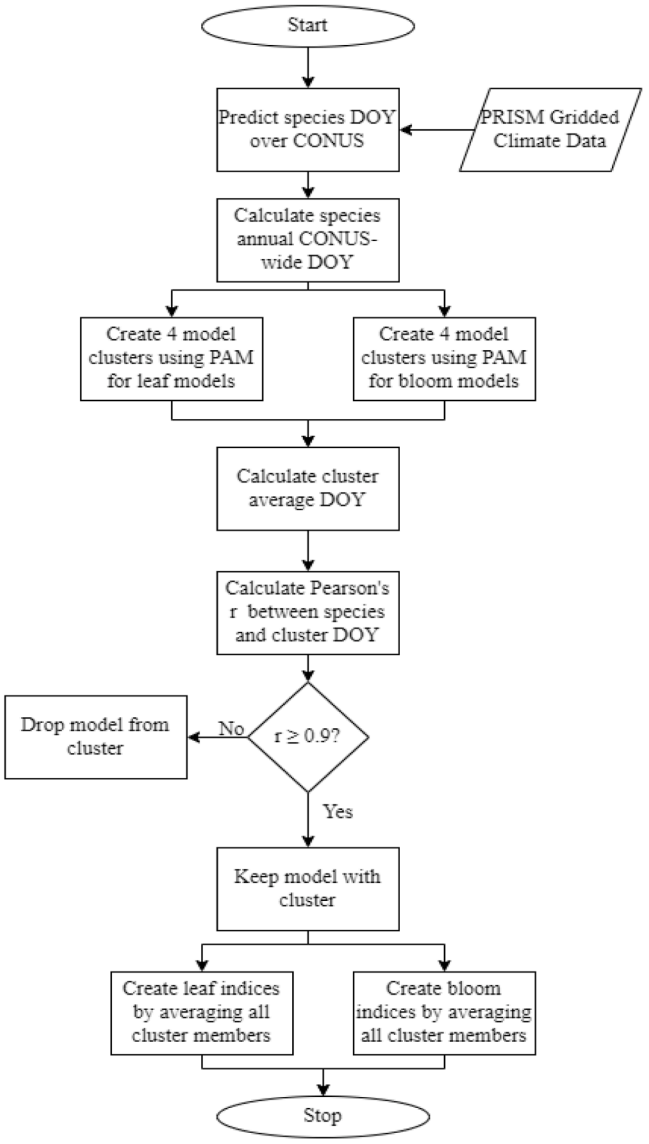


Fig. 4. Flow diagram of the spring development indices (SDIs) construction process.

CONUS and then recalculating the mean absolute error. These error comparisons were completed for the entire CONUS, as well as within each of the nine climate regions (Fig. 2a). Additionally, we calculated the CONUS-wide and regional (for each of the nine climate regions) mean annual phenophase dates for each index (SDI and SI-x) for each year over the comparison/evaluation period (1991–2022).

Our initial comparisons showed that all ten indices (eight SDI + two SI-x) were moderately to highly correlated. We wanted to determine which of these indices were the most different from the others, given that the use of such a large number presents computation issues, as well as likely less useful duplication. Thus, we undertook an ad hoc combination of three exploratory analyses to better understand and discern the separability and degree of independence of the ten indices. First, we visually examined the correlation matrices for the ten indices across the entire CONUS and in each of the nine climate regions. Second, we used principal component analysis (PCA) to probe the indices interrelationships, again for the entire CONUS and in each of the nine climate regions. Our research into PCA suggested that for exploratory analyses, any of the orthogonal rotations are considered comparable (Kim and Mueller 1978). Further, our initial tests showed that the first three components accounted for nearly 99 % of the total variance and were the only ones with eigenvalues greater than one, which is the maximum number that should be retained according to the widely used Kaiser-Guttman criterion. Additionally, the first component was exceptionally dominant (~90 % of total variance) and highly correlated with all indices, suggesting it concentrated the variance all the indices had in common. We hypothesized that examination of the correlations of the individual indices with the second and third components would give an indication of those that likely displayed the most differences. We developed an ad hoc analysis to count the number of times each index was among the top four indices with the highest correlations to these second and third PCA components from two orthogonal rotations (varimax and quartimax) and the unrotated components. Lastly, we employed a third set of tests to assess the statistical difference ($p \leq 0.05$) of the ten indices via ANOVA and post hoc Tukey Honestly Significant Difference (HSD) tests (Tukey 1949).

In order to reveal any trends in the time series, we performed Mann-Kendall trend tests (Mann 1945; Kendall 1975) on each index plus the average January-June and March-May temperatures over the CONUS as well as in each of the nine climate regions during the 1991–2022 period. We also produced detailed graphics for each index showing the linear trends and significance ($p \leq 0.05$) at each grid point (reduced 15 km resolution) over the CONUS during the period to identify internal variations within the regions.

3. Results

3.1. Final species x phenophase models

We produced 41 final species × phenophase models of spring phenology, including leaf models for 21 species (Tables 1 and S1) and bloom models for 20 species (Tables 2 and S2). Among these 41 models, 59 % (24) were geographically dependent, meaning that adding variables that changed by location improved the mean absolute error [MAE] of the models by at least 0.5 days. Two of these models were dependent upon elevation alone, one upon latitude and elevation, seven upon latitude alone, and 14 upon long-term mean (1991–2020) annual temperature. These dependences were incorporated by splitting the model training data into two latitudinal groups for three models, while 20 models used spatial adjustment across the entire CONUS, and one used both split training data and spatial adjustment (model specifications are available in Tables S1 [leaf models] and S2 [bloom models] while the parameters extracted from *phenor* are in Tables S3 [leaf] and S4 [bloom]).

3.2. Spring development indices

The final species × phenophase models were combined into four leaf indices and four bloom indices based on when the events occurred within the spring season: Very Early leaf (VEL), Early leaf (EL), Middle leaf (ML), Late leaf (LL), Very Early bloom (VEB), Early bloom (EB), Middle bloom (MB), and Late bloom (LB, Tables 1 and 2). The best performing SDI (lowest MAE) was MB (MAE = 7.39) while the poorest performing was VEL (MAE = 10.02; Table 1).

3.3. Spring development indices (SDI) performance comparisons

The individual species × phenophase models performed better than all Spring Development Indices (lower MAE) both when the validation data were grouped by index alone (across the entire CONUS, Table 3) and region alone (Tables S5 and S6) as well as when the data were ungrouped. When the data were grouped by both region and index, the SDI performed better than the species × phenophase models for six groups (Tables S5 and S6). The SDIs performed better than SI-x when the validation data were grouped by region alone (Tables S5 and S6) as well as when the data were ungrouped (Table 3) and for all indices except for VEB (Table 3). When the data were grouped by region and temporal

Table 1
Spring Development Indices for leaf models, their component species x phenophse models, mean absolute (MAE), and mean bias (MBE) errors.

Index			*Species × Phenophase Models*		
Index	MAE	MBE	Species Name	MAE	MBE
Very Early leaf	10.02	−1.30	common snowberry (*Symphoricarpos albus*)	13.50	1.54
			northern spicebush (*Lindera benzoin*)	8.51	−2.54
Early leaf	9.48	−0.22	American hornbeam (*Carpinus caroliniana*)	9.69	−0.99
			boxelder (*Acer negundo*)	9.64	1.31
			chokecherry (*Prunus virginiana*)	11.08	0.49
			forsythia (*Forsythia* spp.)	10.06	−0.36
			tuliptree (*Liriodendron tulipifera*)	9.06	−0.42
Middle leaf	9.07	−0.06	black cherry (*Prunus serotina*)	9.75	0.41
			flowering dogwood (*Cornus florida*)	9.00	1.34
			quaking aspen (*Populus tremuloides*)	10.20	−1.23
			red maple (*Acer rubrum*)	9.55	−0.51
			sugar maple (*Acer saccharum*)	8.84	0.43
			sweetgum (*Liquidambar styraciflua*)	6.37	0.38
Late leaf	8.10	0.68	American beech (*Fagus grandifolia*)	8.22	0.07
			eastern cottonwood (*Populus deltoides*)	8.85	0.78
			eastern redbud (*Cercis canadensis*)	9.37	2.35
			green ash (*Fraxinus pennsylvanica*)	7.38	0.47
			northern red oak (*Quercus rubra*)	8.33	−0.24
			striped maple (*Acer pensylvanicum*)	5.73	1.51
			white oak (*Quercus alba*)	8.81	1.44
			yellow birch (*Betula alleghaniensis*)	8.32	0.58

Table 2
Spring Development Indices for bloom models, their component species × phenophase models, mean absolute (MAE), and mean bias (MBE) errors.

Index			*Species × Phenophase Models*		
Index	MAE	MBE	Species Name	MAE	MBE
Very Early bloom	7.96	0.16	beaked hazelnut (*Corylus cornuta*)	5.80	1.11
			forsythia (*Forsythia* spp.)	10.00	0.79
			northern spicebush (*Lindera benzoin*)	7.09	0.16
			red maple (*Acer rubrum*)	8.06	−0.03
Early bloom	8.48	0.85	American hornbeam (*Carpinus caroliniana*)	13.20	1.68
			autumn brilliance serviceberry (*Amelanchier grandiflora-autumn brilliance*)	5.73	0.60
			boxelder (*Acer negundo*)	10.34	0.70
			eastern cottonwood (*Populus deltoides*)	10.16	3.06
			eastern redbud (*Cercis canadensis*)	8.69	2.42
			paper birch (*Betula papyrifera*)	8.76	−2.09
			sugar maple (*Acer saccharum*)	7.27	0.01
Middle bloom	7.39	1.39	American beech (*Fagus grandifolia*)	9.19	−1.07
			chokecherry (*Prunus virginiana*)	7.81	−0.45
			flowering dogwood (*Cornus florida*)	8.13	1.84
			northern red oak (*Quercus rubra*)	6.92	2.1
			striped maple (*Acer pensylvanicum*)	4.92	2.98
			vine maple (*Acer circinatum*)	7.73	−1.00
			white oak (*Quercus alba*)	8.37	1.46
Late bloom	8.86	1.17	black cherry (*Prunus serotina*)	7.59	1.70
			common snowberry (*Symphoricarpos albus*)	11.47	0.08

Table 3
Validation table for each SDI in the conterminous USA. Fields include total number of observations (# Obs), mean absolute error (MAE) for predictions using the species models ($S \times P$), MAE for predictions using the corresponding SDI index, and predictions using leaf or bloom SI-x, as appropriate. Also includes the difference in MAE between the species × phenophase models and SDI as well as the difference between the SDI and SI-x. The final row is the validation for the entire dataset.

		MAE			Δ MAE	
SDI	# Obs	$S \times P$	SDI	SI-x	$S \times P$ – SDI	SDI – SI-x
VEL	158	10.02	11.77	12.72	−1.75	−0.95
EL	627	9.48	10.31	14.23	−0.84	−3.91
ML	2087	9.07	9.85	13.73	−0.78	−3.88
LL	1207	8.10	8.58	12.76	−0.48	−4.18
VEB	942	7.96	10.31	9.98	−2.35	0.33
EB	551	8.48	8.94	9.94	−0.46	−1.00
MB	694	7.39	7.71	8.49	−0.32	−0.78
LB	172	8.86	9.21	12.27	−0.36	−3.05
All SDI	6438	8.55	9.45	12.09	−0.89	−2.65

cohort, SI-x performed better than SDI for six groups (Tables S5 and S6).

Observed interannual variation, as measured by standard deviation, averaged nine days and ranged from five days (ML and MB) to 18 days (VEL). Simulated variation for both the species × phenophase models and Spring Development Indices were smaller (six days for species × phenophase models and seven days for SDI). The SI-x models had the largest variation with an average of 11 days and a maximum variation of 20 days (Fig. 5).

The point-wise simulated grid (CONUS-wide) mean annual date for the leaf indices varied from DOY 78 (2016 Mar. 18; VEL) to DOY 109 (1993 Apr. 19; LL) while the bloom indices varied from DOY 84 (2012 Mar. 24; VEB) to DOY 148 (1993 May 28; LB; Fig. 6). For comparison, the simulated grid point-wise mean annual date for the SI-x Leaf (SIL) varied between DOY 69 (2012 Mar. 9) and DOY 86 (2008 Mar. 26) while SI-x Bloom (SIB) varied between DOY 102 (2012 Apr. 11) and DOY 118 (2008 Apr. 27; Fig. 6).

The regional mean annual date for the leaf indices varied from day 50 (2017 Feb. 19) for VEL in the SE region to day 126 (2011 May 6) for LL in the WNC region while the bloom indices varied from day 58 (2017 Feb. 27) for VEB in the SE region to day 178 (2011 Jun. 27) for LB in the NW region (Figure S1). The standard deviation varied from 3 days (EL in the ENC and WNC regions) to 23 days (VEL in the W region) for the leaf indices and from 2 days (EB in the ENC region) to 28 days (LB in the W and SW regions).

For comparison, the simulated regional mean annual date for SIL varied between day 34 (2017 Feb. 3) in the SE region and day 119 (2008 Apr. 28) in the WNC region while SIB varied between day 68 (2017 Mar. 9) in the SE region and day 157 (2011 Jun. 6) in the NW region (Figure S1). The standard deviation varied from 5 days (in the ENC region) to 38 days (in the W region) for the SIL and 6 days (in the ENC region) to 46 days (in the W region) for SIB.

Across the entire CONUS, species × phenophase models had an average error (MAE) of 8.55 days, while the SDIs ranged from 7.39 to 10.02 days of error, demonstrating that the indices, in general, perform nearly as well as the individual species models at reflecting the timing of activity in their component species. Our indices exhibited an increase of <1 day in error between the SDI (CONUS-wide) and species × phenophase models, demonstrating their utility both as synthetic indices and as proxies for individual species. As a second form of evaluation, we compared SDI performance to that of the SI-x Leaf and Bloom Indices. Overall, SDI predictions averaged more than two days better than SI-x, confirming their increased precision in representing expanded species

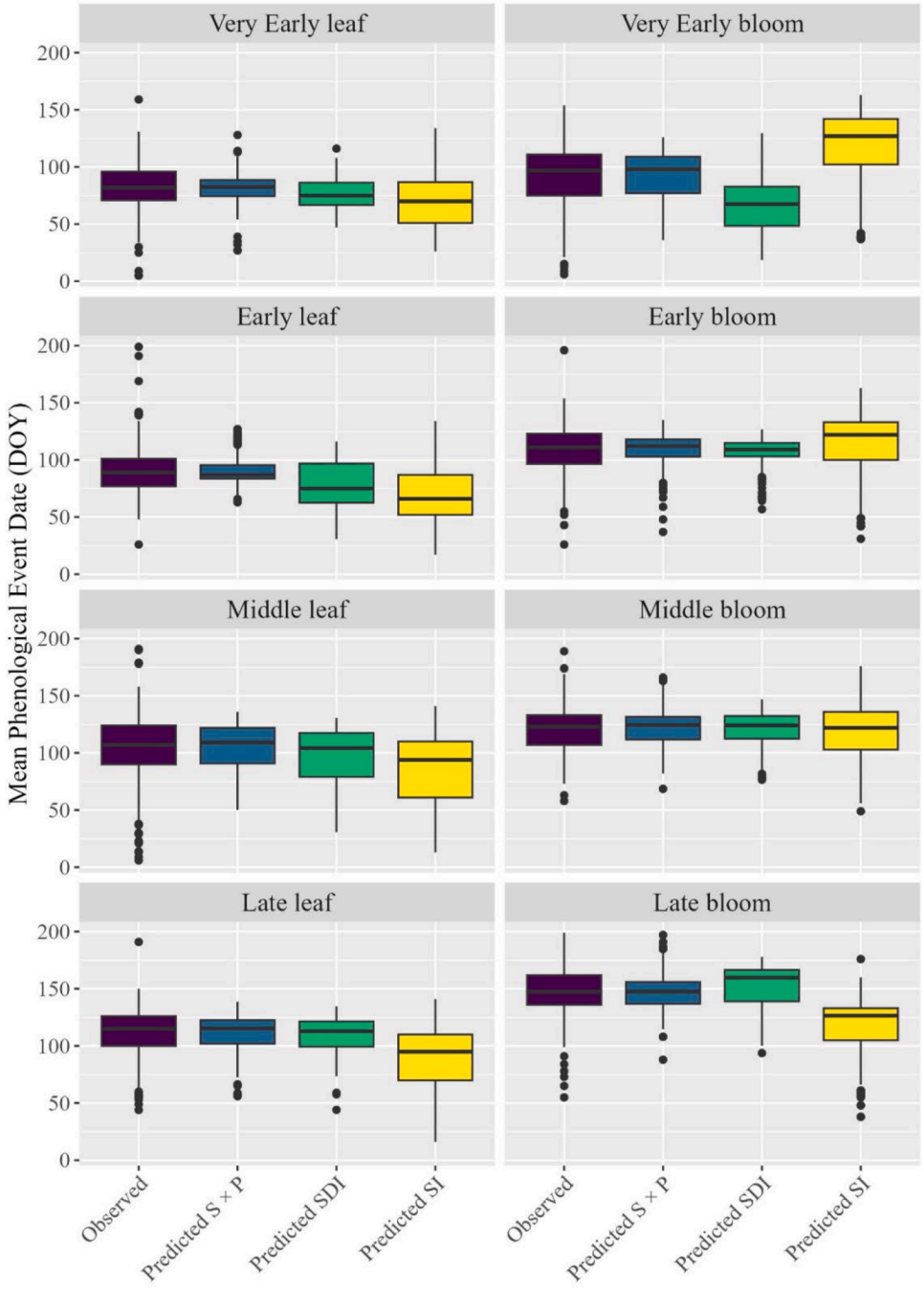


Fig. 5. Boxplot of mean phenophase dates at all plant observation sites. Observed values are compared to the simulated values from the species × phenophase models (Predicted $S \times P$), Spring Development Indices (Predicted SDI), and extended Spring Leaf or Bloom Indices (Predicted SI). Solid lines represent median values while the dots represent outlier data beyond 1.5 x the interquartile range.

phenological responses over a larger portion of the spring season.

Individual species × phenophase models and SDIs exhibited strong performance (70 and 57 percent respectively with MAEs <10 days) across climate regions (Fig. 2a, Tables S5 and S6), though this should be interpreted with caution. As noted earlier, the species are not present in all regions (no observations in the West and West North Central), some are missing from most regions, and only the Northeast region has all SDI component species present. Nevertheless, the results for the climate regions are quite consistent with those for the entire CONUS. Namely, the species × phenophase model's MAE is around a day better than the SDI (range from 0.66 to 1.69) and the SDI MAE is just over two days better than the SI-x (range of 0.81 to 3.53). Not surprisingly, the regions with more of the species present do show lower MAEs overall, indicating regional variations have been able to improve model performance.

In general, individual species models and SDIs generally predicted leaf and bloom phenophasess somewhat earlier than they were observed (Fig. 5). This pattern is not surprising for multiple reasons. First, the observations had higher levels of variation than the model outputs, as it is expected that regression-based models tend to miss some of the extreme variations. Second, observations are expected to show decreasing variation later in the season, due to reductions in temperature extremes and relative strength of the mid-latitude jet stream and associated weather systems (Schwartz and Marotz 1988). This reduction in variation is generally present in our results, although late-bloom SDI

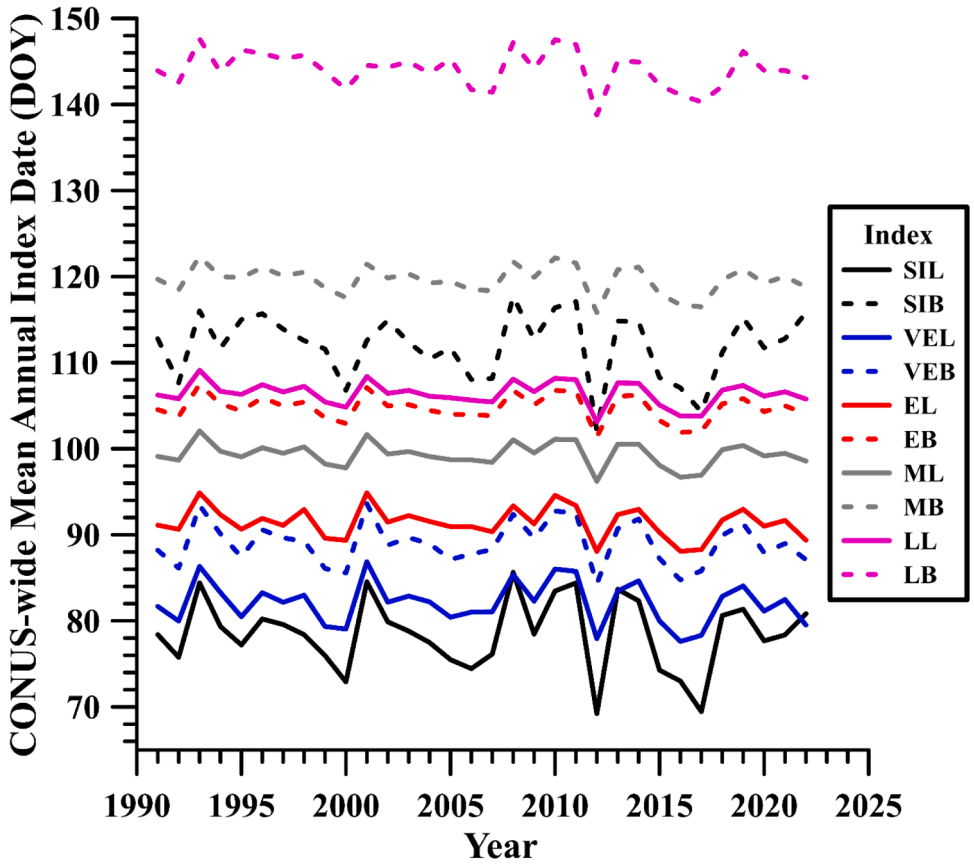


Fig. 6. Mean modeled leaf and bloom dates over the CONUS for 1991 to 2022 for each Spring Development Index (SDI), including the extended Spring Indices Leaf and Bloom (SI-x) for comparison.

is anomalously higher.

The interannual variation in the SI-x models is universally greater than that of SDIs, which is most likely related to the nature of these models. SI-x models are based on cloned plants, which are spatially independent in their responses (i.e., very similar response to environmental variables at all locations), whereas all the SDI component species × phenophase models are based on natural plants, which show adaptation to local conditions. Thus, it is not surprising that SI-x model responses to extreme warm and cold temperatures are greater than those of the species × phenophase models and SDIs, which were constructed using observations from natural plants, which exhibit more genetic diversity, and therefore, more diverse responses to environmental variables. Further, more than half of the species × phenophase models underpinning the SDIs explicitly incorporate spatially dependent modifications in environmental variables.

3.4. *Identifying the most distinct spring development indices*

All indices were highly correlated (Figs. 7a, S2); however, SIL, SIB, and LB consistently exhibited lower correlations with the other indices. Further, the three SDIs VEL, VEB, and EL (though highly correlated with each other) had consistently lower correlations with the previously stated three than any of the remaining indices. From our PCA analyses we found that averaged across the ten indices, the percentage variances explained by the first, second, and third components were 90.4, 6.4, and 1.9 (total of 98.7). The first component concentrated the variance all indices had in common (i.e., highly correlated with all), so examination of the correlations of the individual indices with the second and third components was assumed to give an indication of those most likely to display differences.

Overall, SIL, SIB, LB, and EL showed the highest counts (i.e., the index was in the top four of the PCA, pooled across the unrotated and two rotated analyses) across the entire CONUS, with VEB showing a higher count among the nine regions than EL. Our ANOVA test showed that the CONUS-wide mean annual dates varied between indices (SDI and SI-x, $F = 2006.4$, $p < 0.001$). A post hoc Tukey HSD test showed that all combinations of SDI and SI-x indices had significant differences ($p \leq 0.05$) except for LL and EB ($p = 0.19$, Fig. 7b). Additional ANOVA tests showed that the regional mean annual dates for each region varied across the SDI and SI-x indices ($p \leq 0.05$). A post hoc Tukey HSD test showed that most (378/405) combinations of indices had significant differences ($p \leq 0.05$, Figure S3). ANOVA tests showed that the regional mean annual dates for each SDI and SI-x index varied across the climate regions ($p \leq 0.05$). A post hoc Tukey HSD test showed that most combinations of regions (at least 86 % for each index) showed significant differences ($p \leq 0.05$) in the mean region-wide phenophase date (see Figure S4 and Figure S5). The regions between which the indices varied were regions with differing geographic or climatological features such as the west to east precipitation gradient (e.g. SW compared to C for EL), and mountains (e.g., ENC compared to WNC for VEL).

In summary, the novel SDIs created here show strong correlation with one another as well as with the SI-x indices. For most implementations, utilizing all 10 indices would be excessive. As such, we opted to identify the most distinct indices that might collectively offer the most complete picture of how the full spring season progressed in a particular year at a location. Both SI-x indices (SIL and SIB) are distinct from the new SDIs. Of the SDIs, our comparisons clearly showed that at the CONUS-wide level the Late bloom (LB) and Early leaf (EL) indices were most distinct. At the regional level, Very Early bloom (VEB) also showed noted differences from the other SDIs.

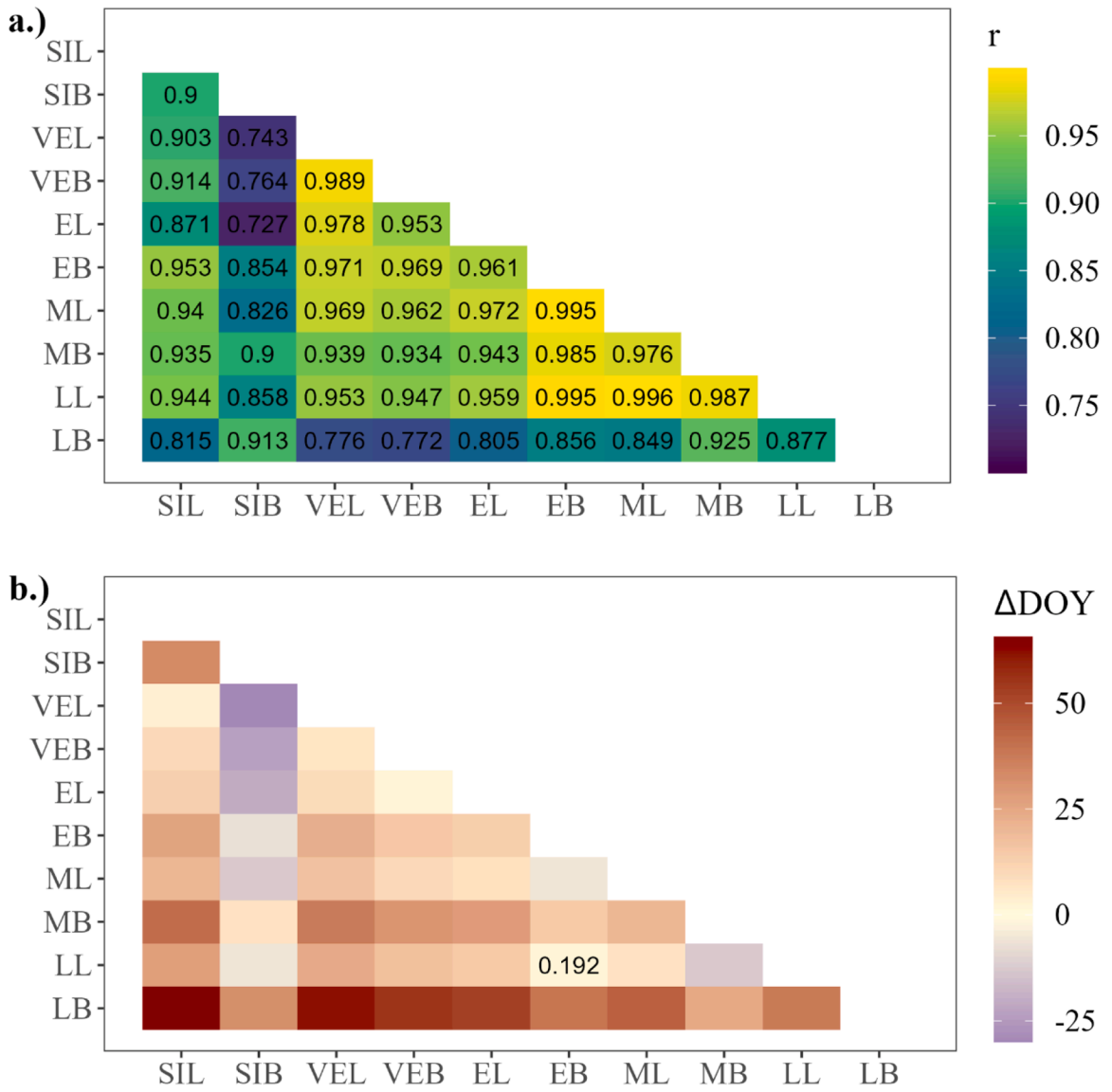


Fig. 7. (a) Pearson's Correlation (r) between mean annual CONUS-wide DOYs for indices and (b) Tukey mean differences between indices (colors) with p-value of difference for boxes that do not have significant differences.

3.5. Evaluating trends in the spring development indices

Mann-Kendall trend tests on each index plus the average January-June and March-May temperatures over the CONUS found no significant trends ($p \leq 0.05$) over 1991–2022. There were also none in each of the nine climate regions, with one exception. There was a weak significant trend (Kendall's tau = 0.26, $p = 0.036$) in the SE region's March-May temperatures. Maps showing the linear trends and significance ($p \leq 0.05$) for each index (15 km resolution) over the CONUS fell into two general patterns. Late bloom (Fig. 8) as well as EL/ML/LL/EB/MB have scattered significant ($p \leq 0.05$) trends mostly in the SE, SW, and W regions (Figure S6). Whereas SI-x bloom (Fig. 9) as well as SIL/VEL/VEB have scattered significant ($p \leq 0.05$) trends mostly in the SW and W regions (Figure S6).

4. Discussion and conclusions

4.1. Major contributions and limitations

In this study, we developed a new suite of synthetic phenological indices designed to predict biological activity at temporal "moments" when quantities of thermal or photo/thermal energy are accumulated to fixed thresholds throughout the spring season and across the CONUS. These new indices complement the widely adopted SI-x (Schwartz et al. 2013a) and thereby can be similarly used to assess how different moments within the season are occurring earlier or later than normal at a location as well as how the timing of these moments is shifting with rapidly changing climate conditions. The premise underpinning of these indices is that they can be used to represent phenological activity in a group of plants. The indices may also be used to reflect activity in insects and other animals that respond to the same sets of environmental conditions (e.g., Gerst et al. 2020).

This study developed leaf and bloom models for woody species that could be projected across the entire CONUS. Unlike the SI-x, which were developed using observations of cloned plants, the new models were constructed using observations of wild and planted individuals of natural (not cloned) plants, which inherently exhibit greater genetic diversity than cloned plants. To accommodate this genetic variation, these models were allowed to be "spatially dependent", meaning their responses to environmental drivers could vary across space. As previously noted, more than half of the models (59 %) exhibited improved performance with the inclusion of spatial variation, underscoring the

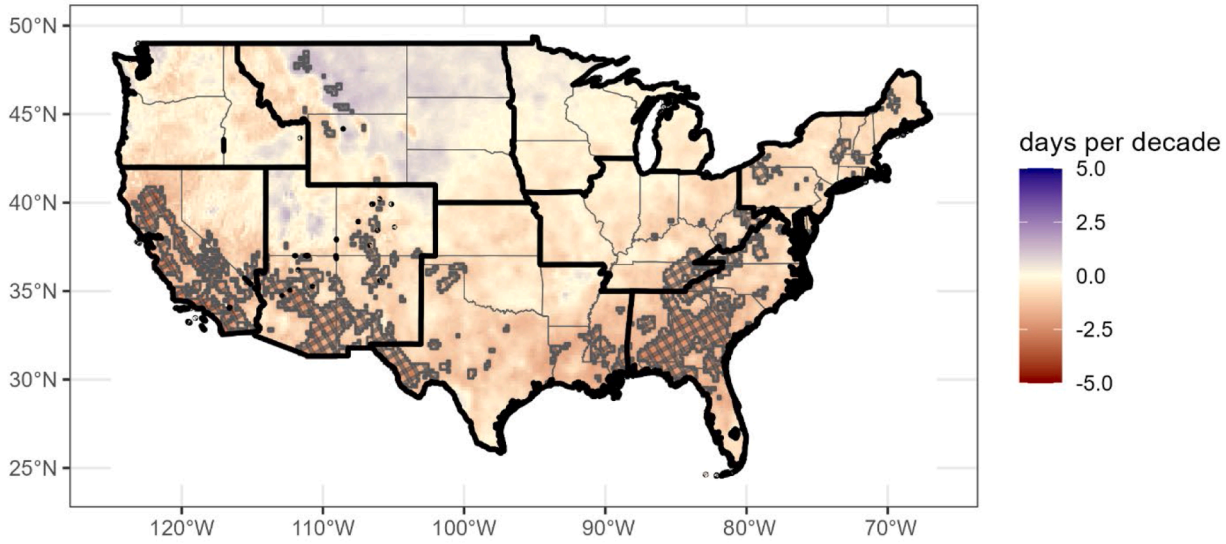


Fig. 8. Trend analysis (linear regression) for Late bloom (LB) index. Background colors correspond to the decadal change in day of bloom. Hatched areas correspond to grid points where the trend is significant ($p \leq 0.05$).

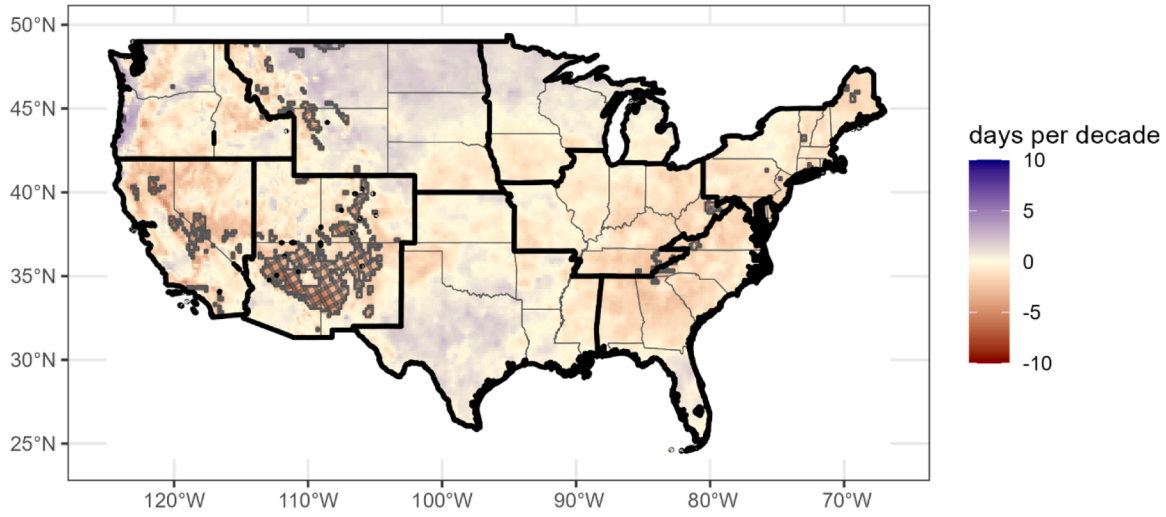


Fig. 9. Trend analysis (linear regression) for extended Spring bloom index (SIB). Background colors correspond to the decadal change in day of bloom. Hatched areas correspond to grid points where the trend is significant ($p \leq 0.05$).

importance of this aspect in phenological models for non-cloned plants. Again, the most incorporated environmental driver with spatial variability was the 1991–2020 site mean annual temperature. This inclusion is unsurprising, since spring phenology is largely shaped by local temperature conditions.

Across the entire CONUS, species × phenophase models had an average error (MAE) of 8.55 days, while the SDIs ranged from 7.39 to 10.02 days of error, demonstrating that the indices, in general, perform nearly as well as the individual species models at reflecting the timing of activity in their component species. These values are in line with what other similar studies report. For example, Kalvāns et al. (2015) reported errors of 2 to 6 days for modeled leaf and bloom phenophasess in silver birch and bird cherry trees in Latvia; Taylor and White (2020) found average errors of 18 to 20 days when developing an ensemble phenology forecasting system using USA-NPN observations, Melaas et al. (2016) used USA-NPN leaf emergence observations to produce models with errors of 6.4 to 10.9 days; Fitzpatrick et al. (2021) developed budburst and leaf out models for a variety of species at one site in Illinois, using USA-NPN data, and found errors of 13 to 23 days and 13 to 19 days respectively; and Crimmins et al. (2017a) found errors ranging from 3.9 (3.5) to 14.1 (18.3) days for leaf (bloom) models of many species when using USA-NPN leaf and bloom observations. In comparison, the generally better performance exhibited by our models likely stems from larger sample sizes and from the inclusion of site-specific information.

The Spring Development Indices showed no widespread significant trends toward an earlier spring onset during the 1991–2022 period, in line with no evidence of significant warming trends in January-June or March-May (except the SE region) over this period in the PRISM temperature data we used. Further, simple regression analyses of average temperature data posted by the U.S. Environmental Protection Agency show the same lack of significant winter and spring trends during our study period (U.S. Environmental Protection Agency 2024). While many studies have shown an overall trend towards an earlier spring: 2.7 to 4.2 days per decade earlier in Europe (Chmielewski and Rötzer 2001; Fu et al. 2014), 5.5 days per decade earlier in China (Ge et al. 2015), and 0.5 to 0.9 days per decade earlier in the US (Piao et al. 2019; Wolfe et al. 2005; Schwartz et al. 2013b;Schwartz and Crimmins 2024) , these studies encompass the mid-1980s, when a global regime shift in

temperature occurred (Reid et al. 2016). There is also a known hiatus in global warming which occurred between 1998 and 2012 (Hu and Fedorov 2017; Medhaug et al. 2017) slowing or stopping any changes in spring phenology (Wang et al. 2019; Su et al. 2022; Yan et al. 2022; Xiong et al. 2023).

4.2. Advice for using models and indices for phenological analyses

The species leaf and bloom models developed in this study were constructed from a database that included at most twelve different years. Further, none of the species' observations extended over the entire CONUS. Still, a few of the species such as red maple exhibited a wide distribution across much of the CONUS, which may compensate to some degree for the short time period. Nevertheless, both limitations should be kept in mind if employing these models to directly predict the phenology of any individual species, especially outside the range of our original data. Importantly, however, that was not our goal. Instead, we produced the species models to combine them into indices, which are not designed to replicate individual species, but broader plant groups. Following Schwartz et al. (2006), our aim was not to reproduce the specific phenology of some types of plants, nor all the details of multi-species phenological data at any site, but instead to process weather data into indices directly related to growth and development of many plants. These indices thus provide baseline assessment of each location's general phenological response over a standard period, supplying a needed context for comparing and evaluating local or regional-scale studies. Thus, SI-x is not most useful as models of lilacs and honeysuckle phenology, but as consistent calculations of accumulated measures of thermal or photo/thermal energy to fixed thresholds (calibrated to plant development stages). For example, indices have been used to predict individual species responses (Gerst et al. 2020), to evaluate the general impacts of climate change on broad groups of plant across the CONUS (Monahan et al. 2016), and to detect modifications in the spring energy balance of the lower atmosphere due to moisture additions from the onset of plant transpiration across many species (Schwartz and Crawford 2001). In summary, indices are specifically designed to be useful for analyses during times and in places where conventional phenological data are not available.

4.3. Applying the spring development indices

The novel Spring Development Indices have strong potential for characterizing a large portion of the spring season, capturing not only whether the start of the season is earlier or later than normal, but also when various moments within the season are met, and whether they are consistently ahead or behind schedule or drift back and forth in response to warm and cool spells. This information can be estimated in real time and can also be calculated for past periods and used to evaluate changes in the timing of spring (e.g., Ault et al. 2015a; Schwartz et al. 2013a). The component 41 species × phenophase models also may have utility for predicting the timing of leaf-out or bloom of individual species, which can be useful in tourism, forestry, and recreation applications as well as ecological and carbon cycle studies. A subset of the new SDIs will be added to the suite of real-time map and short-term forecast products offered by the USA National Phenology Network (Crimmins et al. 2017b). The new indices will eventually also form the basis of long-lead forecasts for the entire CONUS.

4.4. Conclusions

In this study, we established a novel suite of phenological indices. These SDI are comprised of several individual species models that were created using observations of leaf and bloom across the conterminous U. S. These new indices follow the structure and motivation that led to the establishment of SI-x, a pair of widely implemented indices used to characterize springtime plant and animal activity across space and through time. These SDIs represent a greater breadth of the spring season, expanding their utility for both indicating in real time how the season is unfolding, and for characterizing how various "moments" in the spring season have changed through time.

CRediT authorship contribution statement

Joshua J. Hatzis: Methodology, Software, Validation, Formal analysis, Investigation, Data curation, Writing – original draft, Writing – review & editing, Visualization. **Mark D. Schwartz:** Conceptualization, Methodology, Formal analysis, Resources, Writing – original draft, Writing – review & editing, Supervision, Project administration, Funding acquisition. **Toby R. Ault:** Methodology, Writing – review & editing, Funding acquisition. **Alison Donnelly:** Writing – review & editing. **Amanda Gallinat:** Methodology, Software, Data curation, Writing – review & editing. **Xiaolu Li:** Writing – review & editing. **Theresa M. Crimmins:** Conceptualization, Writing – review & editing, Funding acquisition.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability statement

Individual phenometric data and gridded extended Spring Index data are available from the USA National Phenology Network (http://doi.org/10.5066/F78S4N1, accessed on October 1, 2022 and https://doi.org/10.5066/F7XD0ZRK, accessed on May 1, 2023 respectively). Gridded daily atmospheric data are available from the Oregon State University's PRISM Climate Group (https://prism.oregonstate.edu, accessed on May 1, 2023). All source code used for data processing, analysis, and creating figures and tables are available through Open Science Framework (https://osf.io/3wpnq/?view_only=fdb02e3bee3e4db48790f289bd8e949e).

Acknowledgements

Funding was provided by the US National Science Foundation (DEB-2017831, DEB-2017848, DEB-2017815). We are very thankful to the thousands of volunteer participants contributing phenology observations to the USA National Phenology Network's *Nature's Notebook* platform.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.agrformet.2025.110443.

References

Adole, T., Dash, J., Rodriguez-Galiano, V., Atkinson, P.M., 2019. Photoperiod controls vegetation phenology across Africa. Commun. Biol. 2 (1), 391. https://doi.org/10.1038/s42003-019-0636-7.

Allstadt, A.J., Vavrus, S.J., Heglund, P.J., Pidgeon, A.M., Thogmartin, W.E., Radeloff, V. C., 2015. Spring plant phenology and false springs in the conterminous US during the 21st century. Environ. Res. Lett. 10 (10), 104008. https://doi.org/10.1088/1748-9326/10/10/104008.

Ault, T.R., Schwartz, M.D., Zurita-Milla, R., Weltzin, J.F., Betancourt, J.L., 2015a. Trends and natural variability of spring onset in the conterminous United States as evaluated by a new gridded dataset of spring indices. J. Clim. 28 (21), 8363–8378. https://doi.org/10.1175/JCLI-D-14-00736.1.

Ault, T.R., Zurita-Milla, R., Schwartz, M.D., 2015b. A Matlab© toolbox for calculating spring indices from daily meteorological data. Comput. Geosci. 83, 46–53. https://doi.org/10.1016/j.cageo.2015.06.015.

Basler, D., 2016. Evaluating phenological models for the prediction of leaf-out dates in six temperate tree species across central Europe. Agric. For. Meteorol. 217, 10–21. https://doi.org/10.1016/j.agrformet.2015.11.007.

Blümel, K., Chmielewski, F.M., 2012. Shortcomings of classical phenological forcing models and a way to overcome them. Agric. For. Meteorol. 164, 10–19. https://doi.org/10.1016/j.agrformet.2012.05.001.
Chmielewski, F.M., Rötzer, T., 2001. Response of tree phenology to climate change across Europe. Agric. For. Meteorol. 108 (2), 101–112. https://doi.org/10.1016/S0168-1923(01)00233-7.
Chuine, I., Cour, P., Rousseau, D.D., 1999. Selecting models to predict the timing of flowering of temperate trees: implications for tree phenology modelling. Plant Cell Environ. 22 (1), 1–13. https://doi.org/10.1046/j.1365-3040.1999.00395.x.
Chuine, I., Régnière, J., 2017. Process-based models of phenology for plants and animals. Annu Rev. Ecol. Evol. Syst. 48, 159–182. https://doi.org/10.1146/annurev-ecolsys-110316-022706.
Črepinšek, Z., Kajfež-Bogataj, L., Bergant, K., 2006. Modelling of weather variability effect on fitophenology. Ecol. Modell. 194 (1–3), 256–265. https://doi.org/10.1016/j.ecolmodel.2005.10.020.
Crimmins, M.A., Crimmins, T.M., 2019. Does an early spring indicate an early summer? Relationships between intraseasonal growing degree day thresholds. J. Geophys. Res.. 124 (8), 2628–2641. https://doi.org/10.1029/2019JG005297.
Crimmins, T.M., Crimmins, M.A., Gerst, K.L., Rosemartin, A.H., Weltzin, J.F., 2017a. USA National Phenology Network's volunteer-contributed observations yield predictive models of phenological transitions. PLoS. One 12 (8), e0182919. https://doi.org/10.1371/journal.pone.0182919.
Crimmins, T.M., Marsh, R.L., Switzer, J.R., Crimmins, M.A., Gerst, K.L., Rosemartin, A.H., and Weltzin, J.F., 2017b. *USA National Phenology Network gridded products documentation*: U.S. Geological Survey open-file report 2017–1003, 27 p., doi: 10.3133/ofr20171003.
Crimmins, T.M., Gerst, K.L., Huerta, D.G., Marsh, R.L., Posthumus, E.E., Rosemartin, A.H., Switzer, J., Weltzin, J.F., Coop, L., Dietschler, N., Herms, D.A., Limbu, S., Trotter III, R.T, Whitmore, M., 2020. Short-term forecasts of insect phenology inform pest management. Ann. Entomol. Soc. Am. 113 (2), 139–148. https://doi.org/10.1093/aesa/saz026.
Daly, C., Halbleib, M., Smith, J.I., Gibson, W.P., Doggett, M.K., Taylor, G.H., Curtis, J., Pasteris, P.P., 2008. Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. Int. J. Climatol. 28 (15), 2031–2064. https://doi.org/10.1002/joc.1688.
Daly, C., Doggett, M.K., Smith, J.I., Olson, K.V., Halbleib, M.D., Dimcovic, Z., Keon, D., Loiselle, R.A., Steinberg, B., Ryan, A.D., Pancake, C.M., 2021. Challenges in observation-based mapping of daily precipitation across the conterminous United States. J. Atmos. Oceanic Technol. 38 (11), 1979–1992. https://doi.org/10.1175/JTECH-D-21-0054.1.
Denny, E.G., Gerst, K.L., Miller-Rushing, A.J., Tierney, G.L., Crimmins, T.M., Enquist, C.A., Guertin, P., Rosemartin, A.H., Schwartz, M.D., Thomas, K.A., Weltzin, J.F., 2014. Standardized phenology monitoring methods to track plant and animal activity for science and resource management applications. Int. J. Biometeorol. 58 (4), 591–601. https://doi.org/10.1007/s00484-014-0789-5.
Denny, E.G., Crimmins, T.M., 2023. Updates to standardized plant and animal observation protocols of the USA national phenology network. Int. J. Biometeorol. 67, 927–930. https://doi.org/10.1007/s00484-023-02444-0.
Elmendorf, S.C., Jones, K.D., Cook, B.I., Diez, J.M., Enquist, C.A., Hufft, R.A., Jones, M.O., Mazer, S.J., Miller-Rushing, A.J., Moore, D.J., Schwartz, M.D., 2016. The plant phenology monitoring design for the national ecological observatory network. Ecosphere 7 (4), e01303. https://doi.org/10.1002/ecs2.1303.
Fitzpatrick, L., Giambuzzi, P.J., Spreitzer, A., Reidy, B., Still, S.M., Rollinson, C.R., 2021. Improving phenology predictions for sparsely observed species through fusion of botanical collections and citizen-science. Clim. Change Ecol. 2, 100032. https://doi.org/10.1016/j.ecochg.2021.100032.
Forsythe, W.C., Rykiel Jr., E.J., Stahl, R.S., Wu, H., Schoolfield, R.M., 1995. A model comparison for daylength as a function of latitude and day of the year. Ecologic. Model. 80 (1), 87–95. https://doi.org/10.1016/0304-3800(94)00034-F.
Fu, Y.H., Piao, S., Op de Beeck, M., Cong, N., Zhao, H., Zhang, Y., Menzel, A., Janssens, I.A., 2014. Recent spring phenology shifts in western Central Europe based on multiscale observations. Global Ecol. Biogeogr. 23 (11), 1255–1263. https://doi.org/10.1111/geb.12210.
Ge, Q., Wang, H., Rutishauser, T., Dai, J., 2015. Phenological response to climate change in China: a meta-analysis. Global Change Biol 21 (1), 265–274. https://doi.org/10.1111/gcb.12648.
Gerst, K.L., Crimmins, T.M., Posthumus, E.E., Rosemartin, A.H., Schwartz, M.D., 2020. How well do the spring indices predict phenological activity across plant species? Int. J. Biometeorol. 64 (5), 889–901. https://doi.org/10.1007/s00484-020-01879-z.
Govaerts, R., Lughadha, N.E., Black, N., Turner, R., Paton, A., 2021. The world checklist of vascular plants, a continuously updated resource for exploring global plant diversity. Sci. Data 8 (1), 215. https://doi.org/10.1038/s41597-021-00997-6.
Hänninen, H., 1990. Modelling bud dormancy release in trees from cool and temperate regions. Acta Forestalia Fennica 213, 47. https://doi.org/10.14214/aff.7660.
Henniges, Y., Danzeisen, H., Zimmermann, R.D., 2005. Regionale Klimatrends mit hilfe der phänologischen Uhr, dargestellt am Beispiel Rheinland-Pfalz. Umweltwiss. Schadst. Forsch. 17, 28–34. https://doi.org/10.1007/BF03038691.
Hu, S., Fedorov, A.V., 2017. The extreme El Niño of 2015–2016 and the end of global warming hiatus. Geophys. Res. Lett. 44 (8), 3816–3824. https://doi.org/10.1002/2017GL072908.
Hufkens, K., Basler, D., Milliman, T., Melaas, E.K., Richardson, A.D., 2018. An integrated phenology modelling framework in R. Methods Ecol. Evol. 9 (5), 1276–1285. https://doi.org/10.1111/2041-210X.12970.
Joseph, V.R., 2022. Optimal ratio for data splitting. Statistic. Analy. Data Min. 15 (4), 531–538. https://doi.org/10.1002/sam.11583.
Jolly, W.M., Nemani, R., Running, S.W., 2005. A generalized, bioclimatic index to predict foliar phenology in response to climate. Glob. Chang. Biol. 11 (4), 619–632. https://doi.org/10.1111/j.1365-2486.2005.00930.x.
Kalvāns, A., Bitāne, M., Kalvāne, G., 2015. Forecasting plant phenology: evaluating the phenological models for Betula pendula and Padus racemosa spring phases, Latvia. Int. J. Biometeorol. 59, 165–179. https://doi.org/10.1007/s00484-014-0833-5.
Kaufman, L., Rousseeuw, P.J., 1990. Partitioning around medoids (Program PAM). In: Kaufman, L., Rousseeuw, P.J. (Eds.), Finding Groups in Data, pp. 68–125. https://doi.org/10.1002/9780470316801.ch2.
Kendall, M.G., 1975. Rank correlation methods, 4th ed. Charles Griffin and Co. Ltd., London, p. 272.
Kim, J.O., Mueller, C.W., 1978. Introduction to factor analysis: What is it and how to do it. Sage, Beverly Hills, CA.
Kramer, K., 1994. Selecting a model to predict the onset of growth of Fagus sylvatica. J. Appl. Ecol. 172–181. https://doi.org/10.2307/2404609.
Li, X., Ault, T., Evans, C.P., Lehner, F., Carrillo, C.M., Donnelly, A., Crimmins, T., Gallinat, A.S., Schwartz, M.D., 2023. Diverging northern hemisphere trends in meteorological versus ecological indicators of Spring onset in CMIP6. Geophys. Res. Lett. 50 (8), e2023GL102833. https://doi.org/10.1029/2023GL102833.
Liang, L., Schwartz, M.D., 2014. Testing a growth efficiency hypothesis with continental-scale phenological variations for common and cloned plants. Int. J. Biometeorol. 58, 1789–1797. https://doi.org/10.1007/s00484-013-0691-6.
Liang, L., 2019. A spatially explicit modeling analysis of adaptive variation in temperate tree phenology. Agric. For. Meteorol. 266 (267), 73–86. https://doi.org/10.1016/j.agrformet.2018.12.004.
Mann, H.B., 1945. Nonparametric tests against trend. Econometrica 13 (3), 245–259. https://doi.org/10.2307/1907187.
Masle, J., Doussinault, G., Sun, B., 1989. Response of wheat genotypes to temperature and photoperiod in natural conditions. Crop. Sci. 29 (3), 712–721. https://doi.org/10.2135/cropsci1989.0011183X002900030036x.
Medhaug, I., Stolpe, M.B., Fischer, E.M., Knutti, R., 2017. Reconciling controversies about the 'global warming hiatus. Nature 545, 41–47. https://doi.org/10.1038/nature22315.
Melaas, E.K., Friedl, M.A., Richardson, A.D., 2016. Multiscale modeling of spring phenology across deciduous forests in the Eastern United States. Global Change Biol 22 (2), 792–805. https://doi.org/10.1111/gcb.13122.
Monahan, W.B., Rosemartin, A., Gerst, K.L., Fisichelli, N.A., Ault, T., Schwartz, M.D., Gross, J.E., Weltzin, J.F., 2016. Climate change is advancing spring onset across the U.S. National Park system. Ecosphere 7 (10), e01465. https://doi.org/10.1002/ecs2.1465.
Nekovar, J., Koch, E., Kubin, E., Nejedlik, P., Sparks, T., Wielgolaski, F.E., 2008. The history and current status of plant phenology in Europe. European Cooperation in Science and Technology, Brussels, p. 182 pp.
Parameter-elevation Relationships on Independent Slopes Model (PRISM) Climate Group., 2023. PRISM Daily gridded data sets (4km) 10/01/1990-09/30/2022. Northwest Alliance for Computational Science and Engineering: Oregon State University, Corvalis, Oregon, USA. Data set accessed 05/01/2023 at. https://prism.oregonstate.edu.
Park, D.S., Breckheimer, I., Williams, A.C., Law, E., Ellison, A.M., Davis, C.C., 2019. Herbarium specimens reveal substantial and unexpected variation in phenological sensitivity across the eastern United States. Philos. Trans. R. Soc. London 374, 1763. https://doi.org/10.1098/rstb.2017.0394. Ser. B.
Piao, S., Liu, Q., Chen, A., Janssens, I.A., Fu, Y., Dai, J., Liu, L., Lian, X.U., Shen, M., Zhu, X., 2019. Plant phenology and global climate change: current progresses and challenges. Global Change Biol 25 (6), 1922–1940. https://doi.org/10.1111/gcb.14619.
Plants of the World Online (POWO), 2023. "Plants of the world online. Facilitated by the Royal Botanic Gardens, Kew. Published on the Internet. Data accessed on 05/01/2023 at. http://www.plantsoftheworldonline.org/.
R Core Team., 2022. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.
Réaumur, R.A., 1735. Observations du thermomètre faites à Paris pendant l'année 1735, 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 R Des Sci 545–576.
Reid, P.C., Hari, R.E., Beaugrand, G., Livingstone, D.M., Marty, C., Straile, D., Barichivich, J., Goberville, E., Adrian, R., Aono, Y., Brown, R., 2016. Global impacts of the 1980s regime shift. Global Change Biol 22 (2), 682–703. https://doi.org/10.1111/gcb.13106.
Rosemartin, A.H., Denny, E.G., Gerst, K.L., Marsh, R.L., Posthumus, E.E., Crimmins, T.M., and Weltzin, J., 2018. USA national phenology network observational data documentation: U.S. Geological survey open-file report 2018–1060, 24 p., doi: 10.3133/ofr20181060.
Roslin, T., Antão, L., Hällfors, M., Meyke, E., Lo, C., Tikhonov, G., Delgado, M.del M., Gurarie, E., Abadonova, M., Abduraimov, O., Adrianova, O., Akimova, T., Akkiev, M., Ananin, A., Andreeva, E., Andriychuk, N., Antipin, M., Arzamascev, K., Babina, S., Babushkin, M., Ovaskainen, O., 2021. Phenological shifts of abiotic events, producers and consumers across a continent. Nat. Clim. Change 11 (3), 241–248. https://doi.org/10.1038/s41558-020-00967-7.
Schwartz, M.D., 1990. Detecting the onset of spring: a possible application of phenological models. Clim. Res. 1 (1), 23–29. https://www.int-res.com/articles/cr/1/c001p023.pdf.
Schwartz, M.D., 1994. Monitoring global change with phenology: the case of the spring green wave. Int. J. Biometeorol. 38, 18–22. https://doi.org/10.1007/BF01241799.
Schwartz, M.D., 1997. Spring index models: an approach to connecting satellite and surface phenology. In: Lieth, H., Schwartz, M.D. (Eds.), Phenology in seasonal climates. Backhuys Publishers, Leiden, pp. 23–38.

Schwartz, M.D., 1998. Green-wave phenology. Nature 394, 839–840. https://doi.org/10.1038/29670.
Schwartz, M.D., Marotz, G.A., 1988. Synoptic events and spring phenology. Phys. Geogr. 9 (2), 151–161. https://doi.org/10.1080/02723646.1988.10642345.
Schwartz, M.D., Crawford, T.M., 2001. Detecting energy-balance modifications at the onset of spring. Phys. Geogr. 21 (5), 394–409. https://doi.org/10.1080/02723646.2001.10642751.
Schwartz, M.D., Crimmins, T.M., 2024. Phenoclimatology: development and applications in North America. Phys. Geogr. 1–19. https://doi.org/10.1080/02723646.2024.2313783.
Schwartz, M.D., Reiter, B.E., 2000. Changes in North American spring. Int. J. Climatol. 20 (8), 929–932. https://doi.org/10.1002/1097-0088(20000630)20:8<929::AID-JOC557>3.0.CO;2-5.
Schwartz, M.D., Ahas, R., Aasa, A., 2006. Onset of spring starting earlier across the Northern Hemisphere. Global Change Biol 12 (2), 343–351. https://doi.org/10.1111/j.1365-2486.2005.01097.x.
Schwartz, M.D., Betancourt, J.L., Weltzin, J.F., 2012. From Caprio's lilacs to the USA national phenology network. Front. Ecol. Environ. 10 (6), 324–327. https://doi.org/10.1890/110281.
Schwartz, M.D., Ault, T.R., Betancourt, J.L., 2013a. Spring onset variations and trends in the continental United States: past and regional assessment using temperature-based indices. Int. J. Climatol. 33 (13), 2917–2922. https://doi.org/10.1002/joc.3625.
Schwartz, M.D., Beaubien, E.G., Crimmins, T.M., Weltzin, J.F., 2013b. North America. In: Schwartz, M.D. (Ed.), Phenology: an integrative environmental science, 2nd ed. Springer, Dordrecht, pp. 67–89. https://doi.org/10.1007/978-94-007-6925-0.
Stoklosa, J., Daly, C., Foster, S.D., Ashcroft, M.B., Warton, D.I., 2015. A climate of uncertainty: accounting for error in climate variables for species distribution models. Methods Ecol. Evol. 6 (4), 412–423. https://doi.org/10.1111/2041-210X.12217.
Su, M., Huang, X., Xu, Z., Zhu, W., Lin, Z., 2022. A decrease in the daily maximum temperature during global warming hiatus causes a delay in spring phenology in the China–DPRK–Russia cross-border area. Remote Sens. 14 (6), 1462. https://doi.org/10.3390/rs14061462.
Taylor, S.D., White, E.P., 2020. Automated data-intensive forecasting of plant phenology throughout the United States. Ecol. Appl. 30 (1). https://doi.org/10.1002/eap.2025.
Thorndike, R.L., 1953. Who belongs in the family? Psychometrika 18, 267–276. https://doi.org/10.1007/BF02289263.
Tukey, J.W., 1949. Comparing individual means in the analysis of variance. Biometrics 5 (2), 99–114. https://doi.org/10.2307/3001913.
USA National Phenology Network (USA-NPN), 2022. Plant and animal phenology data. Data type: individual phenometrics. 01/01/2010-12/31/2021 for Region: 49.9375°, -66.4791667° (UR); 24.0625°, -125.0208333° (LL). USA-NPN, Tucson, Arizona, USA. Data set accessed 10/01/2022 at doi:10.5066/F78S4N1.
USA National Phenology Network (USA-NPN), 2023a. Spring Indices, historical annual (1981-Previous Year) - first bloom - spring index (Using PRISM Data). 01/01/1991-01/01/2022 for Region: 49.9375°, -66.4791667° (UR); 24.0625°, -125.0208333° (LL). USA-NPN, Tucson, Arizona, USA. Data set accessed 05/01/2023 at doi:10.5066/F7XD0ZRK.
USA National Phenology Network (USA-NPN), 2023b. Spring indices, historical annual (1981-Previous Year) - first leaf - spring index (Using PRISM Data). 01/01/1991-01/01/2022 for Region: 49.9375°, -66.4791667° (UR); 24.0625°, -125.0208333° (LL). USA-NPN, Tucson, Arizona, USA. Data set accessed 05/01/2023 at doi:10.5066/F7XD0ZRK.
USA National Phenology Network (USA-NPN), 2024. Search plants and animals to observe. https://naturesnotebook.usanpn.org/npnapps/species, Accessed 28 October 2024.
U.S. Environmental Protection Agency (EPA). 2024. Climate change indicators: seasonal temperature. Accessed: 2024 March 22 at https://www.epa.gov/climate-indicators/climate-change-indicators-seasonal-temperature.
Wang, J.Y., 1960. A critique of the heat unit approach to plant response studies. Ecology. 41 (4), 785–790. https://doi.org/10.2307/1931815.
Wang, X., Xiao, J., Li, X., Cheng, G., Ma, M., Zhu, G., Altaf Arain, M., Andrew Black, T., Jassal, R.S., 2019. No trends in spring and autumn phenology during the global warming hiatus. Nat. Commun. 10, 2389. https://doi.org/10.1038/s41467-019-10235-8.
Willmott, C.J., Matsuura, K., 2005. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. Climate Res. 30, 79–82. https://doi.org/10.3354/cr030079.
Willmott, C.J., Robeson, S.M., Matsuura, K., 2017. Climate and other models may be more accurate than reported. Eos. 98, 13–14. https://doi.org/10.1029/2017eo074939.
Wolfe, D.W., Schwartz, M.D., Lakso, A.N., Otsuki, Y., Pool, R.M., Shaulis, N.J., 2005. Climate change and shifts in spring phenology of three horticultural woody perennials in northeastern USA. Int. J. Biometeorol. 49, 303–309. https://doi.org/10.1007/s00484-004-0248-9.
Xiong, T., Du, S., Zhang, H., Zhang, X., 2023. Satellite observed reversal in trends of spring phenology in the middle-high latitudes of the Northern Hemisphere during the global warming hiatus. Global Change Biol 29 (8), 2227–2241. https://doi.org/10.1111/gcb.16580.
Yan, Z., Xu, J., Wang, X., Yang, Z., Liu, D., Li, G., Huang, H., 2022. Continued spring phenological advance under global warming hiatus over the Pan-Third Pole. Front. Plant Sci. 13, 1071858. https://doi.org/10.3389/fpls.2022.1071858.