Time to branch out? Application of hierarchical survival models in plant phenology

Sarah C. Elmendorf[a,b,*], Theresa M. Crimmins[c,d], Katharine L. Gerst[c,d], Jake F. Weltzin[c,e]

[a] *Institute for Arctic and Alpine Research (INSTAAR), University of Colorado, Boulder, CO, USA*
[b] *Ecology and Evolutionary Biology Department, University of Colorado, Boulder, CO, USA*
[c] *National Coordinating Office, USA National Phenology Network, Tucson, AZ, USA*
[d] *School of Natural Resources and the Environment, University of Arizona, Tucson, AZ, USA*
[e] *U.S. Geological Survey, Ft. Collins, CO, USA*

ARTICLE INFO

ABSTRACT

The sensitivity of phenology to environmental drivers can vary across geography and species. As such, models developed to predict phenology are typically site- or taxon-specific. Generation of site- and taxon-specific models is limited by the intensive in-situ phenological monitoring effort required to generate sufficient data to parameterize each model. Where in-situ phenological observations exist, the data are often subject to analytical issues due to the limited duration of any individual monitoring program, spotty site- and species- level coverage, lack of standardized methodology, and infrequent or variable census intervals. Together, these characteristics constrain our ability to make phenological inferences outside of select sites and taxa where long-duration, intensive monitoring has occurred. In this study, we leveraged two national, standardized phenology datasets to develop a multi-species and multi-site state-space survival model of the onset of deciduous tree and shrub spring (leaf out) and fall (leaf-color) events across temperate ecoregions of the United States. We used data from two national-scale phenological databases, a 9-year, broadly distributed dataset from the USA National Phenology Network and a 4-year dataset from the National Ecological Observatory Network, to quantify regional and interspecific variation in sensitivity to environmental drivers for both spring and fall leaf phenophases. Spring leaf out was generally promoted by longer days, spring growing degree day accumulation, overwinter chilling, and was suppressed by frost events, whereas fall leaf color was promoted by shorter days and cold accumulation. The sensitivity to most environmental drivers tended to be more variable among species than among the regions as defined here (EPA ecoregions of North America, excluding desert and tropical areas). The results of this study lay the groundwork for incorporating the growing collection of phenological observations into a generalized framework for predicting the transition states for any species, in any location.

1. Introduction

The ability to predict the timing of life cycle events for individual species at given locations is beneficial to many aspects of resource and species management (e.g. optimally timing of land prescribed burns or chemical treatments; (Enquist et al., 2014) and human health (e.g. forecasting the onset or peak of airborne pollen in allergenic taxa (Siljamo et al., 2013). Understanding the drivers to phenology across species also improves our ability to anticipate future changes to ecosystem services as climate conditions change, due the impact of plant phenology on carbon dynamics (Richardson et al., 2013), hydrology (Morisette et al., 2009), and albedo (Williamson et al., 2016).

In recent years, many predictive models have been developed to

elucidate the drivers of plant phenology, ranging from empirical approaches such as regression (Hufft et al., 2018) to more complex process-based models that attempt to explain underlying physiological responses (Basler, 2016). Model construction is frequently data-limited; phenological datasets with both broad taxonomic and geographic coverage that are consistently sampled over decades do not exist. Accordingly, phenology modeling efforts typically fall into one of a few categories: those that analyze one or more species at a single site or small region (Bertelsen's data from the Santa Catalina mountains; the Marshall phenology record from central England, or Thoreau's Concord records (Crimmins et al., 2010; Miller-Rushing and Primack, 2008; Roberts et al., 2015), those that combine phenological observations across different species, regions, or sampling to describe general

patterns using meta-analysis type approaches (e.g Cook et al., 2012; Oberbauer et al., 2013), and those that use space-for-time substitution to parameterize models for single species using a relatively short time-series across multiple sites (Liang and Schwartz, 2014; Melaas et al., 2016; Taylor et al., 2019). Relatively simple phenology models often show equivalent predictive performance with more complex models. However, with all models, transferability of fitted parameters among sites is imperfect (Basler, 2016).

An additional challenge in modeling phenology is that observations often are conducted infrequently because of the effort required in regular censusing, introducing uncertainty into the timing of when an event occurred. Because of this, phenological models have typically been developed separately for the few species or handful of sites with a rich phenological record, which excludes the remaining "long-tail" of phenological observations: those collected over a short duration; those that have few replicates for any species or location; and those that have lower sampling frequency. For phenological studies, we are fortunate that many of these "long-tail" datasets have been collated and publicly archived, making such observations readily available for inclusion in analyses. What remains to be developed is an analytical approach that leverages these data, which can broaden the diversity of species, sites, and time-frames represented, and can enable national-scale inference.

Survival analysis approaches are a promising solution for capitalizing on the long tail of unused phenological data. Survival models are a class of statistical models that are well-suited to accommodating the typical shortcomings of phenology datasets and that have the potential to outperform traditional regression models in predicting phenological events (Templ et al., 2017). Survival analysis has a long history in engineering studies of failure time, as well as in the medical literature in estimating time to death or disease recurrence where observations are right-censored (e.g. Cox, 1972). The analogy to plant monitoring data is clear: Like patients, individual plants may be monitored at long or irregular intervals and in some cases the event of interest occurs after the monitoring period is over. Only a handful of studies have used survival approaches to model climate drivers of plant phenology, despite its potential (e.g. Allen et al., 2014; Diez et al., 2014; Templ et al., 2017; Terres et al., 2013).

We additionally considered the best ways of including data from multiple species and multiple sites. Interspecific variation in phenology dates can be substantial, commonly on the order of several weeks (Lechowicz, 1984). Further, individual species can show different sensitivity to environmental drivers across locations (Ma et al., 2018; Parker et al., 2017; Prevéy et al., 2017; Way and Montgomery, 2015), but see MacKenzie et al. (2018) for a counterexample. However, despite clear differences among species in the dates of phenological transitions, experimental work has shown the phenology of a diversity of deciduous taxa responds similarly to a common set of environmental cues (Flynn and Wolkovich, 2018) and many successful models have been developed by aggregating phenological data using space-for-time substitution. The potential for heterogeneity in sensitivity to environmental drivers among species and regions argues for phenological forcing models that specifically allow for this variability. At the other extreme, modeling each species and region entirely separately dramatically reduces sample sizes. Therefore, to the degree which cues are common across species and regions, these common characteristics should be leveraged to improve models for all species and regions.

Using hierarchical models, it is possible to incorporate variation among species and regions as random effects (Clark, 2007; Bolker, 2008). In this framework, the estimated sensitivity of any individual species or region is influenced by the sensitivities of other taxa and locations, because inferences on species- and region-specific coefficients "borrow strength" from species and regions. Using a multi-species, multi-region approach, the modeled sensitivity of any individual species or region reflects both the sensitivity that would be estimated analyzing each species or region separately and the average sensitivity over all species and regions (Gelman and Hill, 2007). Thus, integrating

survival models into a hierarchical framework further broadens the ability to incorporate additional datasets into the analysis by accommodating observations made on different species and regions.

New, geographically extensive and taxonomically diverse in-situ phenological datasets collected over broad geographic regions broadens possibilities for developing and improving phenological models. The USA National Phenology Network (USA-NPN) and the National Ecological Observatory Network (NEON) are national-scale programs with a primary (USA-NPN) or substantial (NEON) focus on collecting phenological data using concordant, standardized protocols that have substantially increased the availability of in-situ phenological observations across the United States over the past decade (Denny et al., 2014; Elmendorf et al., 2016). The USA-NPN dataset has been used for numerous applications to date, including the identification of climate drivers of phenological transitions for individual species (e.g. Piao et al., 2015; Mazer et al., 2015; Park and Mazer, 2018; Melaas et al., 2016; Taylor et al., 2019), understanding spatial and temporal phenological trends (e.g. Pearse et al., 2017; Liang, 2019), and evaluation of remotely sensed data (e.g. Peng et al., 2017; Zhang et al., 2018). However, none to date have leveraged either the USA-NPN nor the NEON datasets to create hierarchical multi-species models across broad geographic regions.

In this study, we outline a novel analytical approach for analyzing this growing collection of phenology observations across a broad array of ecosystems and species. The objective of our work was to develop hierarchical multi-region, multi-species survival models of the onset of spring (leaf out) and fall (leaf coloration) for over 100 plant species within 15 temperature ecoregions of the U.S. The resulting models demonstrate common drivers of deciduous spring and fall phenology across species in temperate regions of the U.S. and provide estimates of sensitivity for each species and region.

2. Materials and methods

2.1. Model overview

Survival analysis is a modeling approach robust to the challenges that phenology data pose. First, phenology data are essentially time-to-event data, and therefore show different statistical properties than other typical ecological parameters such as size, counts, or concentration. Second, the drivers of phenological transitions vary temporally so that including "average" weather (e.g., April mean daily temperature) as a phenological predictor misrepresents the biology of the system. For example, recent weather events are likely to have an outsized influence on the near-term rate of phenological progression, and weather after a phenophase transition obviously cannot have any effect on the date of the (past) event. Third, in the absence of continuous monitoring, the date on which phenological transitions occur is typically reported with varying degrees of uncertainty, depending on the frequency of observation. Phenological datasets – especially those collected opportunistically or infrequently by volunteers – can suffer from both relatively long periods with no observations as well as variability in temporal frequency of observations. Survival analysis, which can accommodate temporally varying forcing variables and any intercensus interval, is well-suited to data with these characteristics (Templ et al., 2017).

Here, we implement survival modeling using discrete time models where the time-step is a single day, which is typically the smallest intercensus sampling interval for plant phenology. Borrowing from the nest-survival literature (Royle and Dorazio, 2008), we categorize a single plant's spring (or fall) phenology as a state change from one phase (e.g., without leaves) to another (e.g., with leaves). For the purpose of model development, we classify the states as a binary variable that starts in state '1' and has a modeled probability of transitioning to state '0' on each day of year. Mathematically, we can consider this daily probability of changing state as a daily 'hazard' rate, or the complement of the daily 'survival' rate φ_d, which is the probability of

staying in the same (pre-transition) phenophase state on any date.

Thus, on each day (*d*), for each individual (*i*),

$$P(State_{i,d} = 1|State_{i,d-1} = 1) = \varphi_{id} \qquad (1)$$

Once transitioned, the probability of remaining in the terminal state is 100%, i.e.

$$P(State_{i,d} = 0|State_{i,d-1} = 0) = 1 \qquad (2)$$

After transitioning to the (seasonally) terminal state ('0'), reversals are not possible in this framework (i.e., this is a single-season model). The influence of factors that vary at the level of an individual (e.g. geolocation, species, or traits) and/or time period within individual (e.g. daily weather at an individual's location) on phenological timing can be modeled by adding these factors as predictors of the daily survival rates.

We combined data across multiple taxa and regions into a single hierarchical, multi-species, multi-region state-space model separately for spring and for fall phenological events. We used a state-space modeling framework where the daily probability of transitioning from the initial state (leafless for the spring model; lacking colored leaves for the fall model) to the terminal state (with leaves for the spring model; with colored leaves for the fall model) is a function of environmental drivers that occurred in the time prior to each event. The sensitivities of each species and region to each environmental driver were modeled as random effects. Further details on the statistical modeling approach are provided in Appendix A (in Supplementary material).

2.2. Environmental drivers

To assemble a suite of candidate environmental drivers of spring leaf out and fall leaf color to incorporate into multispecies, multi-region models, we surveyed the literature to prioritize the putative drivers (Appendix B in Supplementary material). Based on that analysis, we selected growing degree days (GDD), daylength, chilling, precipitation and frost as environmental forcing variables for the spring leaf model. Environmental forcing variables for the fall leaf color model included cold degree days (CDD), precipitation and daylength. Climate data used to derive these environmental forcing variables were extracted from Daymet (Thornton et al., 2018).

Environmental drivers used as covariates were calculated using an accumulation window approach, reflecting cumulative and dynamic effects on organismal biology. Accumulation windows can be calculated over a moving window period (e.g. the sum of heat over an n-day moving window) or with a fixed start date (e.g. the sum of heat since Jan 1). While the latter is more common, the calendar date selected (usually Jan 1) is fairly arbitrary and a wide range of parameterizations have been used successfully. After considering both the geographic spread of our data and parameterizations used in previous studies, we included as environmental drivers the sum of frost events, cumulative precipitation, and cold degree days (CDD, base 20) over a moving window 14 days to one day prior to each modeled timepoint, the sum of chilling hours using the Utah model (Richardson et al., 1974), over a moving window 135 days to 45 days prior to each modeled timepoint, and growing degree days (GDD, base 0) over a moving window 45 days to one day prior to each modeled timepoint and (Diez et al., 2014; Gill et al., 2015; Gunderson et al., 2012), and the prior day's daylength (Appendix C in Supplementary material). These windows generally created forcing variables for growing degree days, cold degree days, frost, and chill accumulation that were changing unidirectionally over the spring and fall periods. All covariates were scaled to mean 0, unit variance prior to their inclusion in the models to facilitate model interpretation and convergence.

2.3. Filtering and pre-processing of phenological data

We used publicly available datasets from the USA-NPN (USA National Phenology Network, 2018) and NEON (National Ecological Observatory Network, 2018) to predict the onset of leaf out in spring and leaf coloration in fall for deciduous tree and shrub taxa over the temperate regions of the US. The USA-NPN developed standardized monitoring methods for plants and animal taxa that include the assessment of the phenological status (presence or absence) of carefully described, recognizable phenophases (Denny et al., 2014). Citizen science monitoring launched in 2009 and is conducted by tens of thousands of professional scientists and volunteers. NEON is a national-scale ecological observatory which conducts standardized measurements of multiple ecological parameters in 47 terrestrial and 34 aquatics sites across the US. The program has just entered its operational phase, but plant phenology measurements at some sites are available from as early as 2014. USA-NPN and NEON monitoring protocols use identical phenophase definitions and are status-based, such that the raw data on each day of year is an observation of whether and individual is in a particular phenophase (i.e., has leaves) or not (has no leaves) (Denny et al., 2014). At each NEON terrestrial site, plant phenology is at regular intervals by trained technicians. We accessed all available data using the USA-NPN and NEON APIs (application programming interface) and retained observations for species denoted as "Deciduous Trees and Shrubs" by the USA-NPN.

Response variables in the state-space model include, for each monitored individual, in each year, the latest day of year in which an individual was observed to be in the initial state (i.e, leafless for the spring leaf model; having green leaves for the fall color model) and (optionally) the earliest day of year in which an individual was observed to be in the terminal state (with the phenophase "leaves" for the spring leaf model, and having the phenophase "colored leaves" for the fall color model). We converted the raw status observations into 'last observed in initial state (last no)' values and 'first observed in terminal state (first yes)' values based on the annual time-series (Appendix D in Supplementary material). In some cases, absence of the phenophase was not recorded (i.e., no 'last no' values could be extracted from the observations). For these individuals, we assumed the individuals lacked leaves in the middle of winter (Day of Year [DOY] = 1), and had no colored leaves in the middle of summer (DOY = 180). Because the survival analysis approach we took can accommodate any inter-census interval, we did not further filter observations to those with a particular duration between the 'last observed no' and 'first observed yes' values, as is typically necessary for working with phenology data that include some observations with large inter-census intervals.

We assigned each observation based on its location to an EPA level II ecoregion (United States Environmental Protection Agency, 2018), discarding observations that fell into non-temperate (i.e., Mediterranean, tropical and desert-type) regions. In total, we used 16,942 observations (individual*year combinations) representing 128 species in the spring model and 16,244 observations representing 112 species in the fall model (the full list of species included can be found in supplemental tables S1 and S3). Both models covered records across 15 ecoregions of the US (the full list of regions used can be found in supplemental tables S2 and S4) with data from 2009 to 2017 (USA-NPN) or 2014 to 2017 (NEON).

2.4. Model evaluation

Because our data included uncertainty in the response variable (the true date of the phenophase transition may have occurred at any point between the observed last no and observed first yes), we could not rely on traditional methods such as RMSE to assess model fit. Instead, we include a visual assessment of model fit, plotting the model-predicted median phenophase transition DOY vs. observed date range ('last observed no' → 'first observed yes') for each individual*year combination. We also calculated the ranked probability score (RPS) (Epstein, 1969; Murphy, 1970) for a quantitative assessment of relative model fit, following as similar example in Terres et al. (2013). Specifically, we

calculated a relative ranked probability score ($RRPS_i$) for each individual observation, defined as the average of the squared difference of the probability of having transitioned by any day of year (P_{it}) and the actual state (0,1) over all dates prior to and including the 'last observed no' and after and including the 'last observed yes.'

$$RRPS_i = \frac{(\sum_{t=minStart}^{t=lastNo} [P_{i,t}]^2 + \sum_{t=firstYes_i}^{t=maxEnd-1} [1 - P_{i,t}]^2)}{maxEnd - minStart - (firstYes_i - lastNo_i)}$$

where P_{it} is the modeled probability of having transitioned (state = 1) by any given day of year (t). Because the state is an ordered variable and P_{it} is a monotonically increasing function, $RRPS_i$ scores are always higher for individuals with greater measurement uncertainty (i.e., more days between the 'last observed no' and 'first observed yes'). We averaged all $RRPS$ measurements over all observations included in each model for an overall assessment of model quality ($R\bar{R}PS$). $R\bar{R}PS$ scores are only valid when compared to the same input dataset, so, for example, a comparison of the $R\bar{R}PS$ for the spring vs. fall models is not meaningful because the underlying measurement uncertainty of observations included in the two datasets differs. When applied to different models of the same dataset, lower $R\bar{R}PS$ values indicate better model performance.

3. Results

3.1. Model assessment

For the spring leaf out model, we were able to fit a single model including growing degree days (GDD), daylength, chilling, precipitation and frost as environmental forcing variables as well as an intercept (baseline daily probability of remaining in the initial [leafless] state for the spring leaf model [$\hat{R}$ for all parameters < 1.1]). Inclusion of frost days in the fall model created convergence problems. We also experienced a lack of convergence when we included both intercept and daylength as predictors in the same fall model. These issues with convergence were likely due to the large number of days prior to the phenological event that included 0 sum frost days. We therefore fit the fall model without frost as well as alternative forms of the fall model, including either an intercept or daylength but not both. Also, to parallel the fall model, we constructed alternative forms of the spring model with and without an intercept. Spring models that included an intercept showed substantially better fit ($R\bar{R}PS = 0.026$) than those without ($R\bar{R}PS = 0.112$). The fall model that included daylength had a slightly better fit ($R\bar{R}PS = 0.063$) than the model that included an intercept ($R\bar{R}PS = 0.071$).

Visual assessment of both spring and fall models indicated better predictive performance of the spring vs. fall model (Fig. 1). In particular, the fall model erroneously predicted a suite of fall coloration

events to occur very early in the season (before DOY 200). Spring data showed better agreement between predicted and observed dates.

3.2. Sensitivity to environmental drivers

3.2.1. Spring leaf out

On average, spring leaf out was promoted by longer days, higher growing degree day sums, higher amounts of chilling, and was suppressed by recent frost events (Fig. 2A). Precipitation had a slight but nonsignificant suppressing effect on spring leaf out.

Models successfully generated estimates of sensitivity of each taxon and each region to the environmental drivers. With the exception of precipitation, sensitivities of leaf out to environmental drivers varied more among taxa than among regions (Fig. 2C). To illustrate the ability of the models to generate estimates for the large diversity of taxa included here, sensitivities of spring leaf out to one environmental driver (GDD) is presented in Fig. 3. The full suite of parameter estimates for all environmental drivers for each species and region for both the spring models are included in the Supplementary materials (Tables S.1-S.4).

3.2.2. Fall leaf coloration

On average, fall leaf coloration was promoted by short days and higher cold degree day sums. Precipitation had a slight but nonsignificant suppressing effect on fall coloration (Fig. 2B). For fall coloration, we again saw higher variability in sensitivity among species than regions to daylength, but more similar amounts of variability among species and regions for precipitation and cold sensitivity (Fig. 2D). The full suite of parameter estimates for all environmental drivers for each species and region for both the spring models are included in Appendix F.

4. Discussion

Recent growth and standardization of phenology observations across broad spatial scales over multiple years has the potential to enhance our understanding of the drivers of temperate plant phenology. Yet the patchiness of available data results in analytical challenges with respect to spatial, temporal, and taxonomic coverage. We applied a novel analytical framework to address these challenges and better understand the drivers of temperate leaf phenology and quantify sources of variability. On average, spring leaf out was accelerated by longer days, spring growing degree day accumulation, overwinter chilling, and was suppressed by frost events, whereas fall leaf color was accelerated by shorter days and cold accumulation. Comparisons across species and regions indicated climate sensitivity tended to be more variable among species than among regions.
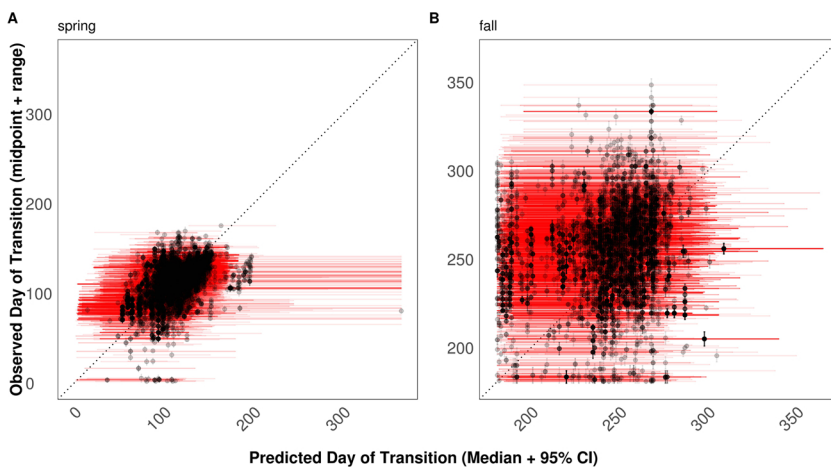

Fig. 1. Predicted versus observed dates of spring leaf out for 128 species (A) and fall leaf coloration for 112 species (B). Figure shows median estimated date of leaf out or fall coloration (x axis, and horizontal 95% CI in red) against observed dates of transition (vertical black bars, lower = 'last observed no,' upper = 'first observed yes,' dot = midpoint). Perfect predictions would fall along the 1:1 line. For purposes of the figure, we included only observations with < 10d intercensus intervals between the 'last observed no' and 'first observed yes,' though the model included all data.
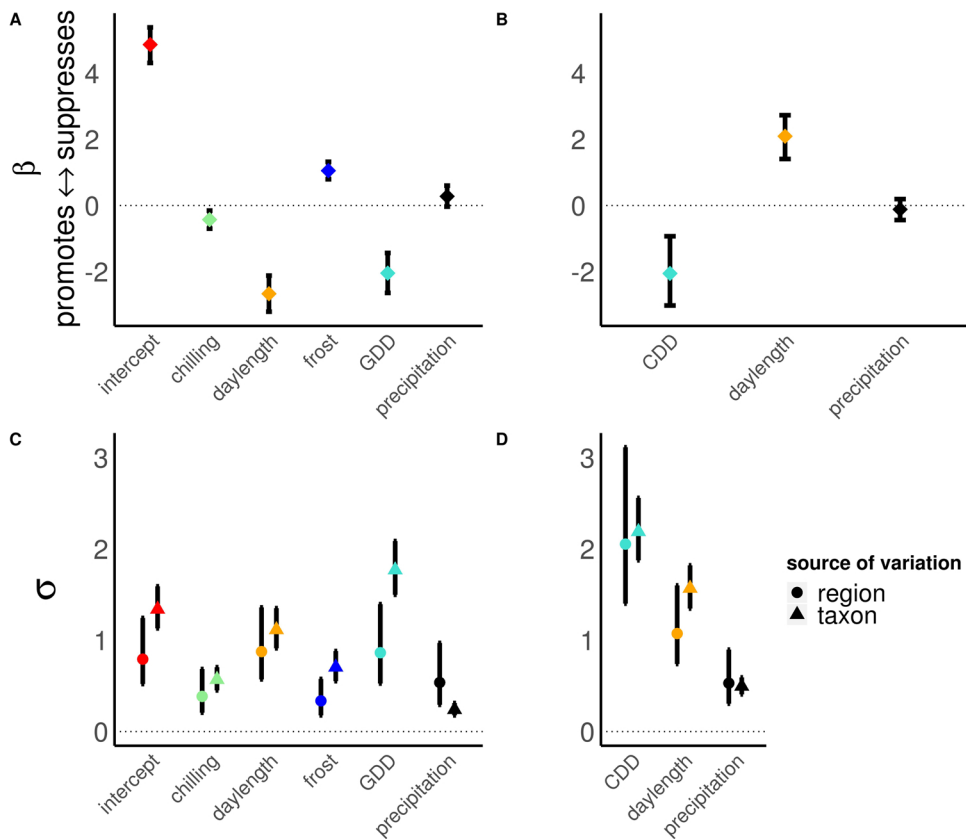
Fig. 2. Mean sensitivities and variance in sensitivities among species and regions of deciduous trees and shrubs phenology to environmental drivers. Panels A (spring leaf out) and B (fall leaf coloration) show estimated mean sensitivity + 95% CI for each environmental driver, plus the baseline hazard rate (intercept). Positive coefficients indicate climate forcing variables that prolong the initial state (i.e., leafless state for spring; green-leaved state for fall), and therefore delay the onset of the phenological transition. Panels C and D show the variability in sensitivity of drivers for spring (C) and fall (D) models among geographic regions (circles) and taxa (triangles).
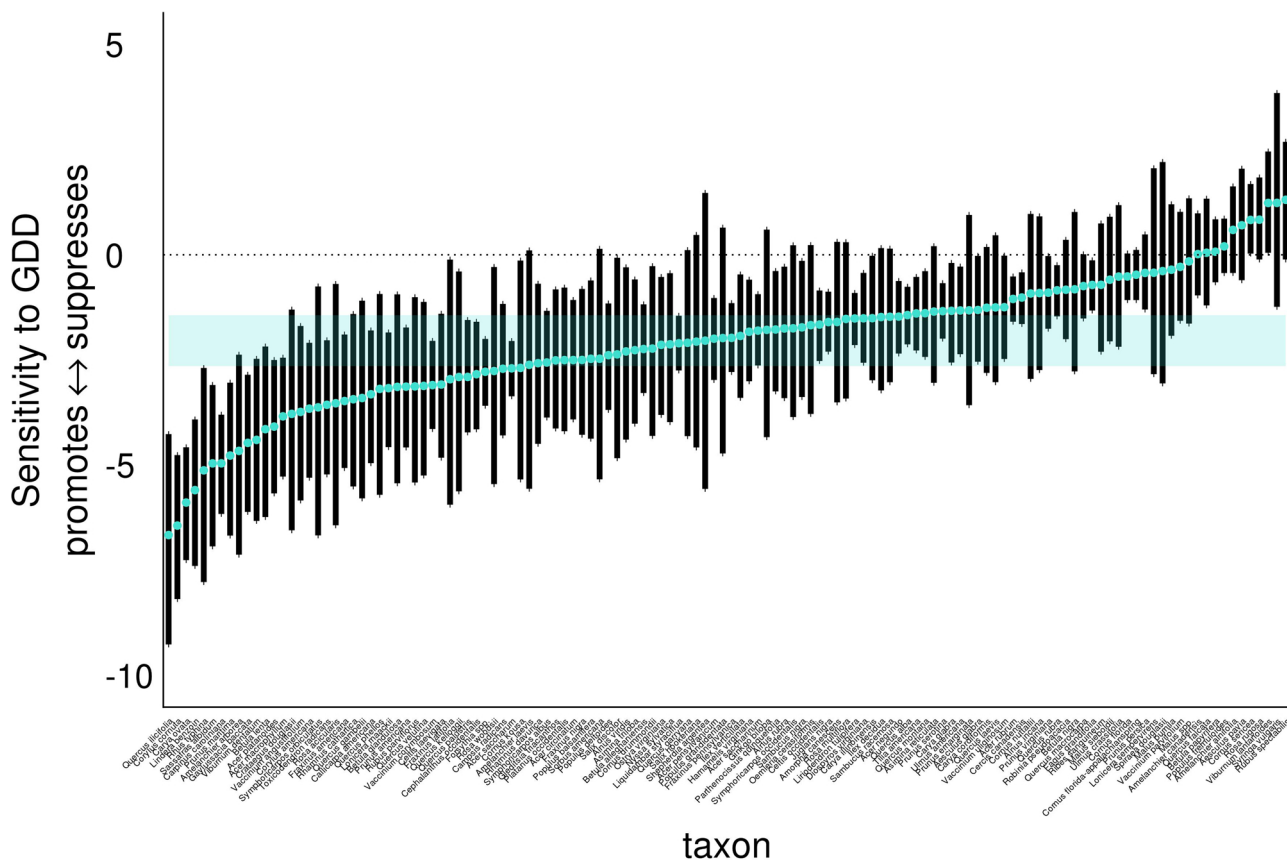
Fig. 3. Sensitivity of spring leaf out to GDD for each taxa (point estimate + 95% CI of each $\beta_{GDD, species}$) term. Negative parameter estimates reflect species where GDD accelerates spring leaf out.

4.1. Advantages of a hierarchical survival modeling approach

We used hierarchical survival models to predict spring and fall phenological events across many taxa and regions. State-space phenology models have been developed using datasets with frequent and regular census intervals (Allen et al., 2014; Diez et al., 2014; Xie et al., 2015). These models can also be easily applied to datasets with variable and irregular observation frequencies. Such flexibility makes state-space approaches particularly well-suited to the data collected by volunteers, where infrequent and unevenly spaced and sampling intervals are the norm. This approach overcomes a common obstacle to embarking on a phenological campaign, namely, a requirement of high-frequency sampling regimes (Fitchett et al., 2015; Menzel, 2002).

In addition to the ability to make use of data with infrequent and/or variable intercensus-intervals, our approach has several advantages over common methods to analyze phenological data in terms of representing our understanding of the biological drivers of phenology. Modeling the daily probability of an event occurring permits us to incorporate weather up to but not beyond the date of a phenological transition as drivers of phenology, which is consistent with our understanding of the biological basis of phenological cues. Using a moving window for forcing accumulation also allows daily probabilities to vary non-monotonically over the growing season; for example, a late spring frost event could suppress the leaf out probability after that event occurs to levels below that before the frost event. Daily models require substantial data volumes to fit; models for only a few species would achieve convergence if each species were modeled separately. At the other extreme, combining datasets across taxa and regions (such as is commonly done in meta-analyses) ignores potential important variability in the sensitivities to environmental drivers among species and regions. Using a hierarchical framework allows us to leverage a large volume of data collected over many sites and species, despite the fact that the dataset as a whole contains relatively few long time series of a single species at a given site. This approach allows us to gain holistic insight into general drivers of deciduous spring and fall phenology across temperate regions of North America, quantify the degree of variability among species and regions, as well as generate estimated phenological sensitivities for over 100 species and 15 ecoregions.

4.2. Drivers of spring and fall phenology

The dominant controls on deciduous tree phenology over most non-desert, temperate regions of the world are generally considered winter chilling, day length and temperature (Korner and Basler, 2010). Our analysis of over 100 deciduous taxa confirmed the importance of these drivers, showing that on average spring leaf out was stimulated by longer days, growing degree sums, and chilling. These results concur with those of Templ et al. (2017), who found a strong positive effect of GDD on flowering using survival modeling, but a weak to insignificant effect of precipitation. The relationship between spring frost or precipitation and spring leaf out events has been less well-studied than that of temperature and daylength. Our results indicated a previously unappreciated role of spring frost in delaying leaf out and a slight trend (different from 0 for a few taxa) of spring precipitation to delay leaf out. The magnitude of the negative (suppressive) effect of frost events was stronger than the positive (promoting) effect of overwinter chilling. The physiological (and ensuing demographic effects) of frost damage are well-documented (Inouye, 2000), so it is unsurprising that plants use recent frost events as a proximal cue for the emergence of sensitive tissues. Our leaf out results parallel a recent synthesis of flowering sensitivity to frost (Park and Mazer, 2018), with both studies highlighting the importance of frost as a driver of the timing of spring phenological events.

The direction of the mean sensitivities of fall leaf coloration to climate drivers was partially consistent with previous studies. Xie et al (2018) found rainfall in autumn cause earlier peak color, while we found a non-significant trend in the same direction. Previous research on the effects of temperature on the timing of fall onset have been conflicting, with evidence to support both advances and delays in leaf coloration with warmer temperatures under natural conditions (see Archetti et al., 2013 and references therein), whereas experimental manipulations tend to suggest a more consistent delaying effect of summer and fall temperature on senescence (Fu et al., 2018; Gunderson et al., 2012). The sign and strength of the relationship between temperature change and onset of fall can vary depending on the time of year and fall phenophase under consideration (Archetti et al., 2013)

We did not detect a strong role of precipitation in the onset of either spring or fall phenological events in this study. In temperate regions such as our study area, plants likely experience moisture stress or abundance through soil moisture, which is a function of soil properties and land cover as well as incoming precipitation (Pan et al., 2003). For example, terrestrial biosphere models such as ORCHIDEE, CLM, LPJ-GUESS and Sheffield- DGVM commonly assume soil moisture physiologically constrains leaf phenology (Abramoff and Finzi, 2015). Improved satellite-based soil moisture products (e.g. Dorigo et al., 2017 or Reichle et al., 2018), when generated on fine-spatial scales for a sufficient number of years, might provide a detectable soil moisture signal beyond the limited trends with precipitation found here.

Species showed consistently strong sensitivity to day length in both the spring (Tables S.1) and fall models (Table S.2). These correlative results are consistent with experiments that show photoperiod sensitivity to be fairly ubiquitous in spring events (Flynn and Wolkovich (2018). The relationship of daylength to fall coloration has not been as thoroughly studied experimentally; however, our results concur with the hypothesis that day length is likely a primary regulator of fall phenology (Way and Montgomery, 2015).

4.3. Variability in sensitivity

Our results demonstrate that sensitivity to environmental drivers varies both by species and by region. The inter-specific and inter-regional variation in sensitivity to environmental drivers found in this study helps explain the heterogeneity of results of previous studies of phenological sensitivity conducted on different species, which has limited transferability of phenological models parameterized at one location to other locations (Basler, 2016). Previous studies have noted that interspecific variability can be as important as climate in predicting landscape level phenological patterns (Xie et al., 2015). Our results support an important role for species composition and location in understanding larger-scale phenological responses. Our process for selecting regions and species for inclusion must be kept in mind when interpreting the magnitude of the region and species variability. We constrained our geographic regions to exclude desert and tropical ecoregions and included species only within a single life form. It is likely that broadening the species pool or geographic locations included would increase the amount of interspecific or inter-regional variability detected. Unsurprisingly, parameter uncertainty declined with sample size, asymptoting at 250-1,000 individuals per species or per region (Appendix E in Supplementary material). This suggests that future data collection activities should focus on species and regions with data densities below that threshold.

4.4. Challenges and opportunities

In this study, predictive capacity for the models was moderate, as evidenced by the fit of the subset of observations with relatively small sampling error to the predicted phenological dates (Fig. 1). Consistent with the literature, our models performed better for spring than fall events (Gallinat et al., 2015). This may reflect a higher degree of variability in the drivers of onset of colored leaves such as disease or other stressors. Future applications using the analytical framework outline here might consider alternative or novel methods for

accumulating phenological forcing variables. For example, modeling forcing variables using critical thresholds rather than using a simple logit-linear relationship; accounting for the varying importance of environmental drivers over different time periods, and/or modeling interactive effects between drivers may yield improved results. An additional consideration is that our estimates of sensitivity to environmental variation combine both sensitivity to interannual climate variability and sensitivity to climate variability over space (due to different locations being sampled in different years). While it is relatively common to combine shorter phenological time series in this way (Liang and Schwartz, 2014; Melaas et al., 2016; Taylor et al., 2019), even within regions there is some evidence that sensitivity to temperature over spatial gradients are stronger than over time (Jochner et al., 2013). As more years of data accumulate from the core-resampled (NEON) datasets here, it may be possible to address the generality of this pattern, and how to best model both spatial and temporal sensitivity. Approaches that account for the relationship between the timing of early-season and later-season phenological events may result in increased performance (Keenan and Richardson, 2015). On the other hand, a similar study (using probit-regression, which has a fairly similar sigmoid form relating linear terms of probability of occurrence as logit-regression) yielded very accurate predictions of flowering date (Allen et al., 2014). This suggests that some of the 'noise' inherent in many distributed phenology datasets, and/or generalizing over such large regions and taxonomically diverse species may have predictive limits.

A limitation of the framework used here is that models were slow to run. Re-running the models over a large combination of forcing variables*periods*functional forms, as has been done for some other efforts (Basler, 2016; Park and Mazer, 2018), is too computationally expensive at present. A more tractable approach to including alternative forms of environmental drivers or accumulation windows might be to use a smaller subset of species and regions for intermodel comparison before refitting the most promising subset on the full dataset.

5. Conclusions

The development of phenological prediction tools requires advancing the field of phenology beyond directional assessment, such as evaluating whether plants generally leaf out earlier with warmer springs, to specific quantitative assessments, such as determining the number of days earlier a species of plant growing in a particular location will leaf out with warming that occurs at particular times of year (Dietze et al., 2018; Houlahan et al., 2017). We demonstrated a novel approach to predicting the timing of phenological events such as leaf out utilizing survival analysis. This hierarchical approach demonstrated strengths over traditional statistical approaches, allowing for the incorporation of data across species and regions and leveraging the large amount of data availability while also offering information on species and region level variability. The results of this study offer a path forward for making widespread phenological predictions by moving from the development of individual models tailored to a single site and taxon to a generalized framework for predicting the transition states for any species, in any location.

Funding

Declaration of Competing Interest

None.

Acknowledgements

Appendices and Supplemental Tables.

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.agrformet.2019.107694.

References

Abramoff, R.Z., Finzi, A.C., 2015. Are above- and below-ground phenology in sync? New Phytol. 205, 1054–1061.

Allen, J.M., Terres, M.A., Katsuki, T., Iwamoto, K., Kobori, H., Higuchi, H., Primack, R.B., Wilson, A.M., Gelfand, A., Silander, J.A., 2014. Modeling daily flowering probabilities: expected impact of climate change on Japanese cherry phenology. Glob. Change Biol. 20, 1251–1263.

Archetti, M., Richardson, A.D., O'Keefe, J., Delpierre, N., 2013. Predicting climate change impacts on the amount and duration of autumn colors in a New England Forest. PLoS One 8, e57373.

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.

Bolker, B.M., 2008. Ecological Models and Data in R. Princeton University Press, Princeton, New Jersey.

Clark, J.S., 2007. Models for Ecological Data: an Introduction. Princeton University Press, Princeton, New Jersey.

Cook, B.I., Wolkovich, E.M., Parmesan, C., 2012. Divergent responses to spring and winter warming drive community level flowering trends. Proc. Natl. Acad. Sci. 109, 9000–9005.

Cox, D.R., 1972. Regression models and life-tables. J Royal Stat Soc Ser B 34, 187–220.

Crimmins, T.M., Crimmins, M.A., Bertelsen, C.D., 2010. Complex responses to climate drivers in onset of spring flowering across a semi-arid elevation gradient. J. Ecol. 98, 1042–1051.

Denny, E.G., Gerst, K.L., Miller-Rushing, A.J., Tierney, G.L., Crimmins, T.M., Enquist, C.A.F., Guertin, P., Rosemartin, A.H., Schwartz, M.D., Thomas, K.A., et al., 2014. Standardized phenology monitoring methods to track plant and animal activity for science and resource management applications. Int. J. Biometeorol. 58, 591–601.

Dietze, M.C., Fox, A., Beck-Johnson, L.M., Betancourt, J.L., Hooten, M.B., Jarnevich, C.S., Keitt, T.H., Kenney, M.A., Laney, C.M., Larsen, L.G., et al., 2018. Iterative near-term ecological forecasting: needs, opportunities, and challenges. Proc. Natl. Acad. Sci. 115, 1424–1432.

Diez, J.M., Ibáñez, I., Silander Jr., J.A., Primack, R., Higuchi, H., Kobori, H., Sen, A., James, T.Y., 2014. Beyond seasonal climate: statistical estimation of phenological responses to weather. Ecol. Appl. 24, 1793–1802.

Dorigo, W.A., Wagner, W., Albergel, C., Albrecht, F., Balsamo, G., Brocca, L., Chung, D., Ertl, M., Forkel, M., Gruber, A., et al., 2017. ESA CCI Soil Moisture for improved Earth system understanding: State-of the art and future directions. Rem. Sens. of Envi. 203, 185–215.

Elmendorf, S.C., Jones, K.D., Cook, B.I., Diez, J.M., Enquist, C.A.F., Hufft, R.A., Jones, M.O., Mazer, S.J., Miller-Rushing, A.J., Moore, D.J.P., et al., 2016. The plant phenology monitoring design for the National Ecological Observatory Network. Ecosphere. n/a-n/a.

Epstein, E.S., 1969. A scoring system for probability forecasts of ranked categories. J. Appl. Meteor. Climatol. 8, 985–987.

Fitchett, J.M., Grab, S.W., Thompson, D.I., 2015. Plant phenology and climate change: progress in methodological approaches and application. Prog. Phys. Geogr. Earth Environ. 39, 460–482.

Flynn, D.F.B., Wolkovich, E.M., 2018. Temperature and photoperiod drive spring phenology across all species in a temperate forest community. New Phytol. 219, 1353–1362.

Fu, Y.H., Piao, S., Delpierre, N., Hao, F., Hänninen, H., Liu, Y., Sun, W., Janssens, I.A., Campioli, M., 2018. Larger temperature response of autumn leaf senescence than spring leaf-out phenology. Glob. Change Biol. 24, 2159–2168.

Gallinat, A.S., Primack, R.B., Wagner, D.L., 2015. Autumn, the neglected season in climate change research. Trends Ecol. Evol. 30, 169–176.

Gelman, A., Hill, J., 2007. Data Analysis Using Regression and Multilevel/Hierarchical Models. Cambridge University Press.

Gill, A.L., Gallinat, A.S., Sanders-DeMott, R., Rigden, A.J., Short Gianotti, D.J., Mantooth, J.A., Templer, P.H., 2015. Changes in autumn senescence in northern hemisphere

deciduous trees: a meta-analysis of autumn phenology studies. Ann. Bot. 116, 875–888.

Gunderson, C.A., Edwards, N.T., Walker, A.V., O'Hara, K.H., Campion, C.M., Hanson, P.J., 2012. Forest phenology and a warmer climate - growing season extension in relation to climatic provenance. Glob. Change Biol. 18, 2008–2025.

Houlahan, J.E., McKinney, S.T., Anderson, T.M., McGill, B.J., 2017. The priority of prediction in ecological understanding. Oikos 126, 1–7.

Hufft, R.A., DePrenger-Levin, M.E., Levy, R.A., Islam, M.B., 2018. Using herbarium specimens to select indicator species for climate change monitoring. Biodivers. Conserv. 27, 1487–1501.

Inouye, D.W., 2000. The ecological and evolutionary significance of frost in the context of climate change. Ecol. Lett. 3, 457–463.

Jochner, S., Caffarra, A., Menzel, A., 2013. Can spatial data substitute temporal data in phenological modelling? A survey using birch flowering. Tree Physiol. 33, 1256–1268.

Keenan, T.F., Richardson, A.D., 2015. The timing of autumn senescence is affected by the timing of spring phenology: implications for predictive models. Glob. Change Biol. 21, 2634–2641.

Korner, C., Basler, D., 2010. Phenology under global warming. Science 327, 1461–1462.

Lechowicz, M.J., 1984. Why do temperate deciduous trees leaf out at different times? Adaptation and ecology of forest communities. Am. Nat. 124, 821–842.

Liang, L., 2019. A spatially explicit modeling analysis of adaptive variation in temperate tree phenology. Agric. For. Meteorol 266–267, 73–86.

Liang, L., Schwartz, M.D., 2014. Testing a growth efficiency hypothesis with continental-scale phenological variations of common and cloned plants. Int. J. Biometeorol. 58, 1789–1797.

Ma, Q., Huang, J.-G., Hänninen, H., Berninger, F., 2018. Reduced geographical variability in spring phenology of temperate trees with recent warming. Agric. For. Meteorol 256–257, 526–533.

MacKenzie, C.M., Primack, R.B., Miller-Rushing, A.J., 2018. Local environment, not local adaptation, drives leaf-out phenology in common gardens along an elevational gradient in Acadia National Park, Maine. Am. J. Bot. 105, 986–995.

Mazer, S.J., Gerst, K.L., Matthews, E.R., Evenden, A., 2015. Species-specific phenological responses to winter temperature and precipitation in a water-limited ecosystem. Ecosphere 6 art98.

Melaas, E.K., Friedl, M.A., Richardson, A.D., 2016. Multiscale modeling of spring phenology across Deciduous Forests in the Eastern United States. Glob. Change Biol. 22, 792–805.

Menzel, A., 2002. Phenology: its importance to the global change community. Clim. Change 54, 379–385.

Miller-Rushing, A.J., Primack, R.B., 2008. Global warming and flowering times in Thoreau's Concord: a community perspective. Ecology 89, 332–341.

Morisette, J.T., Richardson, A.D., Knapp, A.K., Fisher, J.I., Graham, E.A., Abatzoglou, J., Wilson, B.E., Breshears, D.D., Henebry, G.M., Hanes, J.M., et al., 2009. Tracking the rhythm of the seasons in the face of global change: phenological research in the 21st century. Front. Ecol. Environ. 7, 253–260.

Murphy, A.H., 1970. The ranked probability score and the probability score: a comparison. Mon. Weather Rev. 98, 917–924.

National Ecological Observatory Network, 2018. Data Product: NEON.DP1.10055. Provisional data downloaded from http://data.neonscience.org on 2018-07-05. Battelle, Boulder, CO, USA. .

Oberbauer, S., Elmendorf, S., Troxler, T., Hollister, R., Rocha, A., Bret-Harte, M., Dawes, M., Fosaa, A., Henry, G., Høye, T.T., 2013. Phenological response of tundra plants to background climate variation tested using the International Tundra Experiment. Philos. Trans. R. Soc. B Biol. Sci 368, 20120481.

Pan, F., Peters-Lidard, C.D., Sale, M.J., 2003. An analytical method for predicting surface soil moisture from rainfall observations. Water Resour. Res. 39.

Park, I.W., Mazer, S.J., 2018. Overlooked climate parameters best predict flowering onset: assessing phenological models using the elastic net. Glob. Change Biol. 24, 5972–5984.

Parker, T.C., Tang, J., Clark, M.B., Moody, M.M., Fetcher, N., 2017. Ecotypic differences in the phenology of the tundra species Eriophorum vaginatum reflect sites of origin. Ecol. Evol. 7, 9775–9786.

Pearse, W.D., Davis, C.C., Inouye, D.W., Primack, R.B., Davies, T.J., 2017. A statistical estimator for determining the limits of contemporary and historic phenology. Nat. Ecol. Evol. 1, 1876–1882.

Peng, D., Wu, C., Li, C., Zhang, X., Liu, Z., Ye, H., Luo, S., Liu, X., Hu, Y., Fang, B., 2017. Spring green-up phenology products derived from MODIS NDVI and EVI: inter-comparison, interpretation and validation using National Phenology Network and AmeriFlux observations. Ecol. Indic. 77, 323–336.

Piao, S., Tan, J., Chen, A., Fu, Y.H., Ciais, P., Liu, Q., Janssens, I.A., Vicca, S., Zeng, Z., Jeong, S.-J., et al., 2015. Leaf onset in the northern hemisphere triggered by daytime temperature. Nat. Commun. 6, 6911.

Prevéy, J., Vellend, M., Rüger, N., Hollister, R.D., Bjorkman, A.D., Myers-Smith, I.H., Elmendorf, S.C., Clark, K., Cooper, E.J., Elberling, B., et al., 2017. Greater temperature sensitivity of plant phenology at colder sites: implications for convergence across northern latitudes. Glob. Change Biol. 23, 2660–2671.

Reichle, R., De Lannoy, G., Koster, R.D., Crow, W.T., Kimball, J.S., Liu, Q., 2018. SMAP L4 Global 3-hourly 9 Km EASE-Grid Surface and Root Zone Soil Moisture Analysis Update, Version 4. NASA National Snow and Ice Data Center Distributed Active Archive Center, Boulder, Colorado USA. https://doi.org/10.5067/60HB8VIP2T8W.

Richardson, A.D., Keenan, T.F., Migliavacca, M., Ryu, Y., Sonnentag, O., Toomey, M., 2013. Climate change, phenology, and phenological control of vegetation feedbacks to the climate system. Agric. For. Meteorol. 169, 156–173.

Richardson, E.A., Seeley, S.D., Walker, D.R., 1974. A model for estimating the completion of rest for "Redhaven" and "Elberta" peach trees. HortScience 9, 331–332.

Roberts, A.M.I., Tansey, C., Smithers, R.J., Phillimore, A.B., 2015. Predicting a change in the order of spring phenology in temperate forests. Glob. Change Biol. 21, 2603–2611.

Royle, J.A., Dorazio, R.M., 2008. Modeling survival. Hierarchical Modeling and Inference in Ecology: the Analysis of Data From Populations, Metapopulations and Communities. Academic Press, pp. 347–377.

Siljamo, P., Sofiev, M., Filatova, E., Grewling, Ł., Jäger, S., Khoreva, E., Linkosalo, T., Ortega Jimenez, S., Ranta, H., Rantio-Lehtimäki, A., et al., 2013. A numerical model of birch pollen emission and dispersion in the atmosphere. Model evaluation and sensitivity analysis. Int. J. Biometeorol. 57, 125–136.

Taylor, S.D., Meiners, J.M., Riemer, K., Orr, M.C., White, E.P., 2019. Comparison of large-scale citizen science data and long-term study data for phenology modeling. Ecology 100, e02568.

Templ, B., Fleck, S., Templ, M., 2017. Change of plant phenophases explained by survival modeling. Int. J. Biometeorol. 61, 881–889.

Terres, M.A., Gelfand, A.E., Allen, J.M., Silander, J.A., 2013. Analyzing first flowering event data using survival models with space and time-varying covariates. Environmetrics 24, 317–331.

Thornton, P.E., Thornton, M.M., Mayer, B.W., Wei, Y., Devarakonda, R., Vose, R.S., Cook, R.B., 2018. Daymet: Daily Surface Weather Data on a 1-km Grid for North America, Version 3. Data set accessed 5 July 2018. ORNL DAAC Oak Ridge Tenn. USA. https://doi.org/10.3334/ORNLDAAC/1328.

United States Environmental Protection Agency, 2018. Ecoregions of North America. USA. Data set accessed 2018-07-05 at. https://www.epa.gov/eco-research/ecoregions-north-america.

USA National Phenology Network, 2018. Plant and Animal Phenology Data. Data type: Status and Intensity (2009-2017). NPN, Tucson, Arizona, USA. Data set accessed 2018-07-05 at. https://doi.org/10.5066/F78S4N1V.

Way, D.A., Montgomery, R.A., 2015. Photoperiod constraints on tree phenology, performance and migration in a warming world. Plant Cell Environ. 38, 1725–1736.

Williamson, S.N., Barrio, I.C., Hik, D.S., Gamon, J.A., 2016. Phenology and species determine growing-season albedo increase at the altitudinal limit of shrub growth in the sub-Arctic. Glob. Change Biol. 22, 3621–3631.

Xie, Y., Ahmed, K.F., Allen, J.M., Wilson, A.M., Silander, J.A., 2015. Green-up of deciduous forest communities of northeastern North America in response to climate variation and climate change. Landsc. Ecol. 30, 109–123.

Xie, Y., Civco, D.L., Silander, J.A., 2018. Species-specific spring and autumn leaf phenology captured by time-lapse digital cameras. Ecosphere 9, e02089.

Zhang, X., Liu, L., Liu, Y., Jayavelu, S., Wang, J., Moon, M., Henebry, G.M., Friedl, M.A., Schaaf, C.B., 2018. Generation and evaluation of the VIIRS land surface phenology product. Remote Sens. Environ. 216, 212–229.