

ECOGRAPHY

Research

Herbarium specimens provide reliable estimates of phenological responses to climate at unparalleled taxonomic and spatiotemporal scales

Tadeo H. Ramirez-Parada, Isaac W. Park and Susan J. Mazer

T. H. Ramirez-Parada (<https://orcid.org/0000-0002-9900-8373>) ✉ (tadeo@ucsb.edu), I. W. Park and S. J. Mazer, Dept of Ecology, Evolution and Marine Biology, Univ. of California, Santa Barbara, CA, USA.

Ecography

2022: e06173

doi: 10.1111/ecog.06173

Subject Editor: Jorge Soberon

Editor-in-Chief: Miguel Araújo

Accepted 15 June 2022



Understanding the effects of climate change on the phenological structure of plant communities will require measuring variation in sensitivity among thousands of co-occurring species across regions. Herbarium collections provide vast resources with which to do this, but may also exhibit biases as sources of phenological data. Despite general recognition of these caveats, validation of herbarium-based estimates of phenological sensitivity against estimates obtained using field observations remains rare and limited in scope. Here, we leveraged extensive datasets of herbarium specimens and of field observations from the USA National Phenology Network for 21 species in the United States and, for each species, compared herbarium- and field-based estimates of peak flowering dates expected under standardized temperature conditions, and of sensitivity of peak flowering time to geographic and interannual variation in mean minimum temperatures (TMIN). We found strong agreement between herbarium- and field-based estimates for standardized peak flowering time ($r=0.91$, $p < 0.001$) and for the direction and magnitude of sensitivity to both geographic TMIN variation ($r=0.88$, $p < 0.001$) and interannual TMIN variation ($r=0.82$, $p < 0.001$). This agreement was robust to substantial differences between datasets in 1) the long-term TMIN conditions observed among collection and phenological monitoring sites and 2) the interannual TMIN conditions observed in the time periods encompassed by both datasets for most species. Our results show that herbarium-based sensitivity estimates are reliable among species spanning a wide diversity of life histories and biomes, demonstrating their utility in a broad range of ecological contexts, and underscoring the potential of herbarium collections to enable phenoclimatic analysis at taxonomic and spatiotemporal scales not yet captured by observational data.

Keywords: citizen science, natural history collections, spatiotemporal climate variation, specimen-based research, temperature sensitivity, validation studies



www.ecography.org

© 2022 The Authors. Ecography published by John Wiley & Sons Ltd on behalf of Nordic Society Oikos

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

Introduction

Widespread shifts in plant phenology (i.e. the timing of life cycle events) due to climate change have the potential to significantly alter species distributions (Chuine 2010), trophic interactions (Renner and Zohner 2018), species persistence (Cleland et al. 2012) and community structure (Miller-Rushing et al. 2008). A trend towards earlier flowering and leaf out in response to warming has demonstrated that phenology is highly sensitive to climate variation, but sensitivity varies widely among regions and taxa (Cook et al. 2012, Park 2014, Menzel et al. 2020), and even within species (Song et al. 2020, Love and Mazer 2021, Pearson et al. 2021), limiting our ability to extrapolate documented patterns to unstudied systems. Therefore, predicting plant phenological responses to climate change and their impact across communities, landscapes and biomes will require significant increases to the geographic and taxonomic coverage of phenoclimatic analysis.

Regular field observations of individual plants allow precise records of the date of phenological events and are the gold-standard for the study of phenology–climate relationships. However, observational datasets spanning enough time to permit detection of phenological shifts are scarce and predominantly consist of phenological records from North America and Western Europe (Cook et al. 2012, Templ et al. 2018), limiting their utility in assessing phenology–climate relationships across many unstudied taxa and biomes (Wolkovich et al. 2014, Tang et al. 2016). Moreover, field-based time series of phenology are usually available only at single sites for most species, constraining estimation of phenological responses to climate to small subsets of their ranges.

In contrast, herbarium specimens capture snapshots of the reproductive status of individual plants in space and time, and with hundreds of millions of records worldwide increasingly available digitally, provide unique opportunities to expand the taxonomic and spatiotemporal coverage of phenoclimatic studies (Willis et al. 2017, Meineke et al. 2018). In recent years, researchers have leveraged specimens to study phenology–climate relationships (Jones and Daehler 2018, Heberling et al. 2019), estimating phenological responsiveness for thousands of species (Park and Mazer 2018) and generating results qualitatively consistent with those from field studies (Calinger et al. 2013). However, potential biases in collection practices could yield inaccurate estimates of a species' phenology and its sensitivity to climate. For example, while field observations can pinpoint the timing of a phenological event with known degrees of uncertainty, herbarium specimens may have been collected anytime between the onset and termination of a phenophase, or botanists may preferentially collect individuals in specific phenophases (e.g. peak flowering, Panchen et al. 2019), potentially compromising collection dates as reliable proxies for the dates of phenological events, especially the onset and termination of a phenophase. Additionally, the opportunistic collection of specimens could result in sampling of early or late flowering individuals that may not accurately reflect the phenological behavior of their populations.

Despite these caveats, studies designed to validate herbarium-based estimates of phenology and its sensitivity to climate using field observations are few and limited in scope. Most validation studies have been restricted to areas with long records of field observations and specimen collections covering a small portion of species' ranges (Miller-Rushing et al. 2006, Robbirt et al. 2011, Davis et al. 2015). In turn, the only studies comparing herbarium- and field-based phenological records at large spatial scales have not aimed to validate phenological sensitivity estimates (Spellman and Mulder 2016, Park and Mazer 2018). Some studies have compared herbarium- versus field-based estimates of sensitivity for a single species (Robbirt et al. 2011), or conducted pooled, multi-species analyses that do not enable validation of estimates for individual species (Miller-Rushing et al. 2006, Park 2012). As an exception, Davis et al. (2015) used herbarium and field data for 20 species collected in Middlesex County (Massachusetts, USA), finding overall agreement between data for the direction of phenological responses to spring temperature variation; however, sensitivity estimates derived from the two sources tended to differ in magnitude and were not positively correlated among species. Collectively, these studies have shown herbarium specimens are promising data sources for phenoclimatic analysis, but their limited scope and the mismatch in estimates between data types in Davis et al. (2015) make it difficult to establish to what extent specimens may represent generally valid resources for the study of phenology–climate relationships.

In this study, we provide a multi-species comparison of herbarium- and field-based estimates of peak flowering sensitivity to spatiotemporal variation in mean minimum temperatures during the months preceding the mean flowering date of each species (TMIN). We used two geographically extensive datasets obtained from herbaria across the United States and from field observations aggregated by the USA National Phenology Network (USA-NPN, hereafter, 'NPN'; Schwartz et al. 2012). These data included a total of 21 species spanning diverse life histories and biomes and included phenological observations across thousands of unique site-year combinations throughout the United States. These data substantially exceed the sample sizes of previous validation studies and enabled us to compare herbarium- and field-based estimates of sensitivity to climate variation over both space and time and across a broad range of ecological contexts.

We measured peak flowering time sensitivities to both geographic and interannual variation in TMIN because, among conspecifics distributed across large geographic scales, associations between phenology and climate might be driven both by phenotypic plasticity and by local adaptation to long-term climatic conditions among populations (Anderson et al. 2012). While associations between phenology and interannual climate variation are thought to predominantly reflect plastic responses, correlations between phenology and long-term, mean climatic conditions over space may be strongly influenced by local adaptation across populations (Delgado et al. 2020). To the extent that phenology–climate relationships over space and time have different drivers, they

may also differ in magnitude or direction. Therefore, we leveraged the spatiotemporal scale of these datasets to partition observed variation in temperature across sites and years into interannual and geographic components, comparing herbarium- and field-based estimates of sensitivity to both sources of temperature variation in all examined species. In doing so, this study provides the first concurrent validation of herbarium-based estimates of phenological sensitivity to spatial and temporal variation in temperature.

Material and methods

Phenological data

Field observations consisted of all records of flowering onset and termination available in the NPN database, representing an initial 1 105 764 phenological observations. To ensure data quality, we retained only observations for which flowering onset and termination dates had an arbitrary maximum error of 14 d. Accordingly, we filtered the data to include only records for which the date of flowering onset was preceded by an observation of the same individual without flowers no more than 14 d prior, and for which the flowering termination date was followed by an observation of the same individual without flowers no more than 14 d later. The remaining field observations had an average maximum error of 6.4 d for flowering onset, and of 6.6 d for flowering termination.

Herbarium data consisted of an initial 894 392 digital specimen records archived by 72 herbaria across North America (see Supporting information for a list). We removed all specimens not explicitly recorded as being in flower, or for which GPS coordinates or dates of collection were not available. We further filtered both datasets by only retaining species that were found in both datasets and that were represented by observations at an arbitrary minimum of 15 unique sites in both datasets. To better align the geographic range of each dataset for each species, we filtered herbarium observations to include only specimens within the range of latitudes and longitudes represented among field observations in the NPN data. Finally, we retained only species represented by 70 or more herbarium specimens to ensure sufficient sample sizes for phenoclimatic modeling (Park and Mazer 2018). This procedure identified a final set of 21 native species represented in 3243 field observations across 1406 unique site–year combinations, and a final sample of 5405 herbarium specimens across 4906 unique site–year combinations (Fig. 1). These species represented 15 families and 17 genera, spanning a diverse range of life history strategies and growth forms, including evergreen and deciduous shrubs and trees, as well as herbaceous perennials and annuals. Our focal species covered a wide variety of biomes and regions including western deserts, Mediterranean shrublands, oak woodlands and Eastern deciduous forests (Table 1).

We employed the day of year of collection (henceforth ‘DOY’) of each specimen collected while in flower as a proxy of flowering dates. Flowering specimens could have been

collected at any point between onset and termination, and botanists may preferentially collect individuals at their flowering peak for many species (Panchen et al. 2019). Therefore, specimen DOYs are more likely to reflect peak flowering dates than onset or termination dates (Primack et al. 2004). To increase the phenological equivalence of field and herbarium observations, we used the median date between flowering onset and termination for each observation in the NPN data as a proxy for peak flowering time. Median flowering dates also had a maximum error of 14 days, with an average maximum error among observations of 6.5 days. Because flowering spanned year ends for some species (e.g. *Quercus agrifolia*), we accounted for the artificial DOY discontinuity between 31 December (DOY = 365–366) and 1 January (DOY = 1) by converting DOY into a circular variable using an Azimuthal correction (Park and Mazer 2018).

Climate data

Daily minimum temperatures mediate key developmental processes including the break of dormancy, floral induction and anthesis (Reeves and Coupland 2000). Therefore, we used minimum surface temperatures averaged over the three months leading up to (and including) the mean flowering month for each species (hereafter ‘TMIN’) as the climatic correlate of flowering time in this study; consequently, the specific months over which temperatures were averaged varied among species. Using TMIN calculated over different time periods instead (e.g. during spring for all species) did not qualitatively affect our results. Then, we partitioned variation among sites into spatial and temporal components, characterizing TMIN for each observation by the long-term mean TMIN at its site of collection (henceforth ‘TMIN normals’), and by the deviation between its TMIN in the year of collection (for the three-month window of interest) and its long-term mean TMIN (henceforth ‘TMIN anomalies’) (Supporting information; see Munson and Long 2017 for an example of this approach).

For each site, we obtained a monthly time series of TMIN from January 1901 to December 2016, using ClimateNA ver. 6.30 (Wang et al. 2016), a software package that interpolates 4 km² resolution climate data from PRISM (PRISM Climate Group, Oregon State Univ., <<http://prism.oregon-state.edu>>) to generate elevation-adjusted climate estimates. To calculate TMIN normals, we averaged observed TMIN for the three months leading up to the mean flowering date of each species across all years between 1901 and 2016 for each site. TMIN anomalies relative to long-term conditions were calculated by subtracting TMIN normals from observed TMIN conditions in the year of collection. Therefore, positive and negative values of the anomalies respectively reflect warmer-than-average and colder-than-average conditions in a given year (Supporting information).

Pooling across species, herbarium records showed slightly cooler TMIN normals than did NPN field observations, and spanned a wider envelope encompassing warmer and cooler long-term conditions in the months leading up

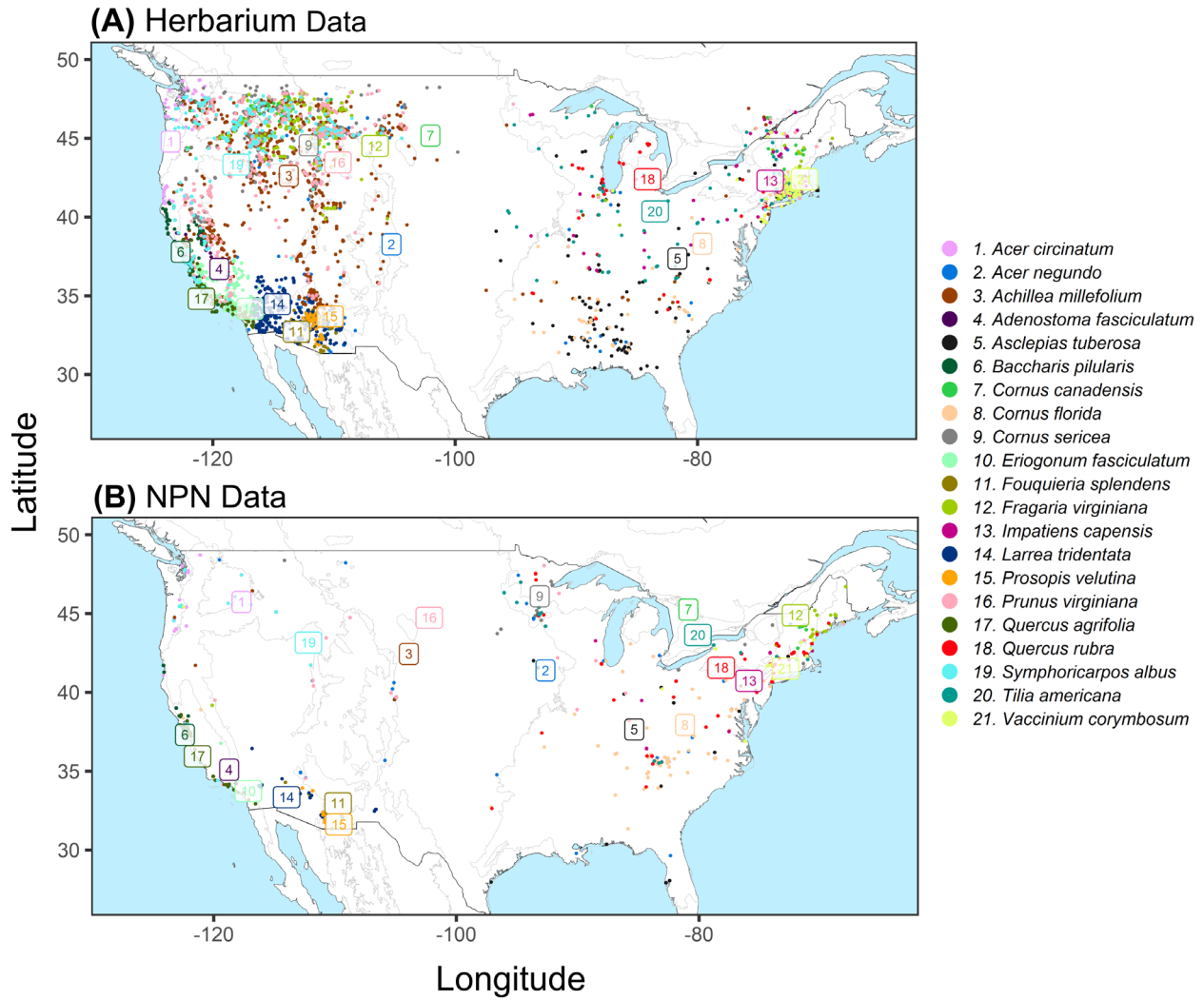


Figure 1. Geographic distribution of herbarium specimens (A) and field observations from the USA National Phenology Network (USA-NPN) (B) for 21 species in the continental United States. Numbered labels represent the centroid of the spatial distribution of observations for each species, obtained by calculating the average latitude and longitude among data points for each species.

to mean flowering dates (Fig. 2A). Specimen collection dates spanned a long period (1901–2016) largely preceding the onset of rapid warming trends, while NPN observations were all conducted in recent years (2009–2020). Consequently, TMIN anomalies in the NPN dataset encompassed warmer conditions than those in the herbarium dataset, both globally and for most species (Fig. 2A, C). Among species, differences between datasets in the width and median of TMIN normal and anomaly envelopes varied substantially (Fig. 3B, C), but relative differences in TMIN envelopes among species was largely consistent in both datasets (Supporting information).

Analyses

We compared estimates of sensitivity to spatiotemporal variation in TMIN derived from herbarium specimens and field observations, concurrently measuring the effects of TMIN normals and anomalies on peak flowering time for

each species-by-dataset combination. We combined herbarium and field records in a single dataset, which we analyzed using a varying-intercepts, varying-slopes Bayesian mixed-effect model. The model fitted species-specific intercepts and slopes and treated them as random effects stemming from community-level distributions (defined by separate ‘hyperparameters’) for field and herbarium records. This hierarchical structure improves estimation of parameters for species with low sample sizes by using community-level information and estimates from better-sampled species. In turn, the Bayesian inference framework enables direct measurement of uncertainty for all parameters.

We used peak flowering DOY for each observation i in the combined dataset as a response, which was assumed to be normally distributed, with mean μ_i and species-specific standard deviation σ_i (Eq. 1):

$$\text{DOY}_i \sim N(\mu_i, \sigma_i) \quad (1)$$

Table 1. Life history, ecoregions of phenological monitoring, estimates of standardized flowering time and of sensitivities to mean minimum temperature (TMIN) normal and anomaly obtained from herbarium (H.) and USA National Phenology Network (USA-NPN) observations and absolute differences ($|\Delta|$) between data types for these estimates among 21 species in the continental United States (see 'Methods' section, Eq. 1–3). Ecoregion codes are for Level I Ecoregions under the United States Environmental Protection Agency (US EPA) classification system (Omernik and Griffith 2014), and correspond to the top two ecoregions where most of the phenological observations for each species (USA-NPN and herbarium combined) were conducted. Where only one ecoregion code is provided, a majority of phenological observations occurred within a single ecoregion, with the second most frequent ecoregion accounting for < 10% of total observations for that species. DOY = day of year of collection

Species	Life history	Ecoregion	Sample size (unique locations)		Standardized flowering DOY		Sens. to TMIN normal (d °C ⁻¹)		Sens. to TMIN anomaly (d °C ⁻¹)	
			H.	NPN	H.	NPN	H.	NPN	H.	NPN
<i>Acer circinatum</i>	Deciduous shrub	7.0, 6.0	77(71)	50(27)	154	120	34	-2.4	0.5	-2.5
<i>Acer negundo</i>	Deciduous tree	8.0, 13.0	149(135)	134(42)	126	100	26	-2.0	1.8	-4.8
<i>Achillea millefolium</i>	Perennial herb, rhizomatous	6.0, 10.0	1368(1323)	43(18)	193	192	1	-4.1	2.8	-3.5
<i>Adenostoma fasciculatum</i>	Evergreen shrub	11.0	106(103)	95(17)	154	165	11	-4.9	2.7	-4.6
<i>Asclepias tuberosa</i>	Perennial herb, tuberous	8.0	132(125)	51(23)	223	187	36	-4.0	0.5	-0.5
<i>Baccharis pilularis</i>	Evergreen shrub	11.0, 8.0	144(143)	359(59)	259	237	22	4.4	0.7	2.2
<i>Cornus canadensis</i>	Perennial herb, rhizomatous	5.0, 6.0	128(124)	100(28)	160	158	2	-4.6	0.3	-3.8
<i>Cornus florida</i>	Deciduous tree or shrub	8.0	118(113)	321(108)	97	90	7	-4.3	0.1	-3.7
<i>Cornus sericea</i>	Deciduous shrub	6.0, 7.0	369(358)	92(24)	165	159	6	-3.1	0.3	-1.9
<i>Eriogonum fasciculatum</i>	Evergreen shrub	11.0, 10.0	440(420)	279(39)	174	176	2	-7.4	2.9	-1.7
<i>Fouquieria splendens</i>	Deciduous shrub	10.0, 12.0	107(98)	79(19)	111	111	0	-5.1	1.3	-3.9
<i>Fragaria virginiana</i>	Perennial herb, stolonate	6.0, 8.0	371(360)	152(27)	145	140	5	-4.5	0.8	-3.9
<i>Impatiens capensis</i>	Annual herb	8.0, 5.0	136(131)	74(15)	209	213	4	2.4	0.8	4.1
<i>Larrea tridentata</i>	Evergreen shrub	10.0	433(399)	320(32)	114	121	7	-10.5	0.3	-1.8
<i>Prosopis velutina</i>	Evergreen tree	10.0	74(69)	150(15)	153	164	11	-7.5	1.2	-4.0
<i>Prunus virginiana</i>	Deciduous tree	6.0, 10.0	551(530)	111(25)	151	138	13	-2.5	3.2	-4.2
<i>Quercus agrifolia</i>	Evergreen tree	11.0	96(96)	181(37)	97	97	0	-2.8	1.8	-2.9
<i>Quercus rubra</i>	Deciduous tree	8.0, 5.0	94(84)	382(69)	183	119	64	-2.2	0.7	-0.5
<i>Symphoricarpos albus</i>	Deciduous shrub	6.0, 11.0	281(275)	117(15)	193	190	3	-1.7	0.2	-2.4
<i>Tilia americana</i>	Deciduous tree	8.0, 5.0	93(88)	62(20)	204	187	17	-1.7	0.1	0.4

Ecoregion codes: 5.0 = Northern Forests, 6.0 = Northwestern Forested Mountains, 7.0 = Marine West Coast Forests, 8.0 = Eastern Temperate Forests, 10.0 = North American Deserts, 11.0 = Mediterranean California, 12.0 = Southern Semiarid Highlands, 13.0 = Temperate Sierras.

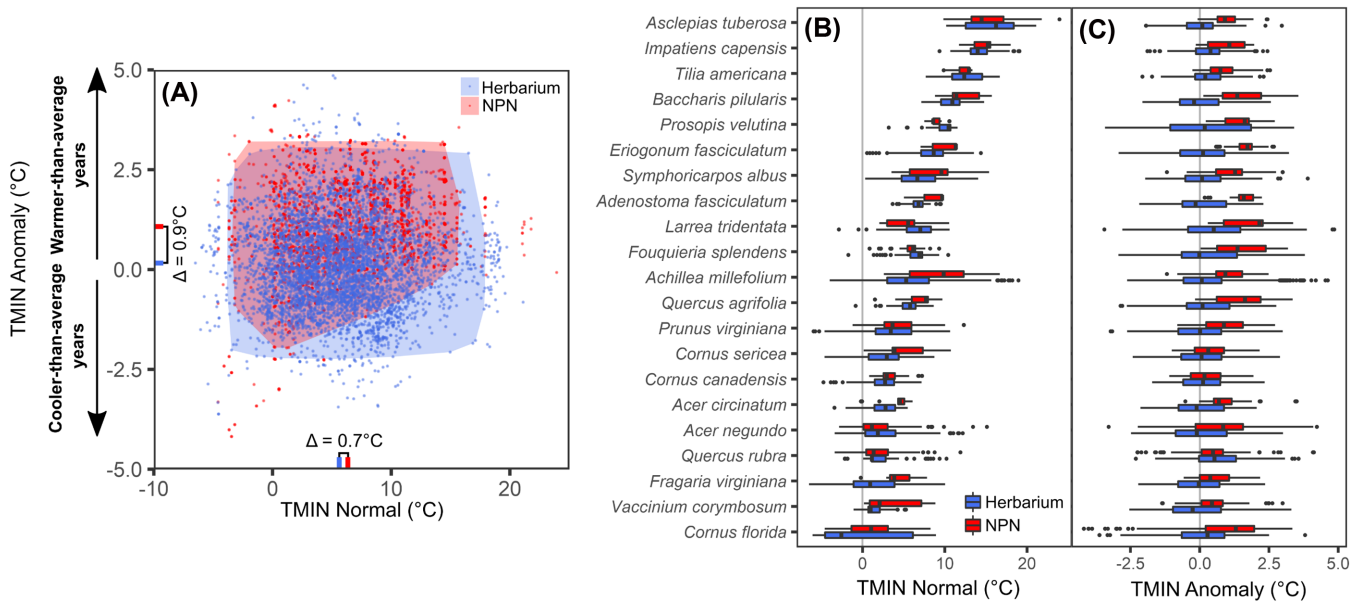


Figure 2. Envelopes for the central 90% of mean minimum temperature (TMIN) normal and TMIN anomaly records associated to herbarium specimens and field observations for all species pooled (A), and by species (B, C). The small red and blue segments located on the x- and y-axes in (A) indicate the mean values of each climatic variable for herbarium (blue) versus field (red) data.

We modeled μ_i as a linear function of TMIN normal (TMIN Norm.), and TMIN anomaly (TMIN Anom.) for each observation i . To obtain intercepts and slopes unique to each species-by-dataset combination, we used two dummy variables (with values of 0 or 1) respectively indicating whether each observation was obtained from field observations in the NPN (F_i) or from herbarium records (H_i). This resulted in the inclusion of only NPN or herbarium observations when a given parameter was estimated (i.e. model terms were turned ‘on and off’ depending on data type). For each data type, the model yielded species-specific intercepts representing standardized flowering dates expected under mean TMIN normal and mean TMIN anomaly conditions (herbarium: α_1 ; NPN: α_2), species-specific sensitivities (i.e. regression slopes) for TMIN normal (herbarium: β_1 ; NPN: β_2) and species-specific sensitivities for TMIN anomaly (herbarium: β_3 ; NPN: β_4) (Eq. 2):

$$\begin{aligned} \mu_i = & \alpha_1 \times H_i + \alpha_2 \times F_i + \beta_1 \times \text{TMIN Norm}_i \times H_i \\ & + \beta_2 \times \text{TMIN Norm}_i \times F_i + \beta_3 \times \text{TMIN Anom}_i \times H_i \\ & + \beta_4 \times \text{TMIN Anom}_i \times F_i \end{aligned} \quad (2)$$

To account for co-variation among parameters, we assumed that community-level distributions for intercepts and slopes were generated by a multivariate normal distribution with a vector of hyper-means μ and a variance–covariance matrix Σ (Eq. 3):

$$(\alpha_1, \alpha_2, \beta_1, \beta_2, \beta_3, \beta_4) \sim N(\mu, \Sigma) \quad (3)$$

The diagonals in Σ correspond to community-level variances for each intercept and slope, whereas off-diagonal values correspond to the covariances between parameters among species.

Priors in the model were weakly informative, with wide, 0 centered normal distributions for intercepts, slopes and rate parameters for exponential distributions (used to obtain species-specific variances). For the variance–covariance matrix Σ , we used a Lewandowski–Kurowicka–Joe (LKJ) Cholesky covariance prior, with $\eta=1$ to allow for high correlations among parameters. Posterior distributions were obtained using Hamiltonian Monte Carlo (HMC) in Stan (code provided in the Supporting information), implemented in R Studio ver. 1.4.1106 using the ‘rstan’ package ver. 2.21.2 (<www.r-project.org>, Stan Development Team 2020). We implemented a non-centered parameterization to improve sampling of the parameter space. Sampling was done using two MCMC chains with training and sampling periods of 1000 iterations each. All parameters had Gelman–Rubin statistics (‘R-hat’) values close to 1, and visual examination of trace plots confirmed convergence. Bulk and tail effective sample size were both high relative to the total number of samples.

To evaluate the correlation between herbarium- and field-derived estimates accounting for differences in the number of sampled sites between datasets, we calculated weighted Pearson correlation coefficients between maximum a posteriori (MAP) estimates for field data and herbarium specimens generated by the model, using the minimum number of unique sites in the NPN or the herbarium dataset for each species as weights. Alternative weighting schemes (e.g. using total sample sizes instead) yielded nearly identical results.

We assessed whether, among species, mismatches between herbarium- versus field-based estimates could be explained

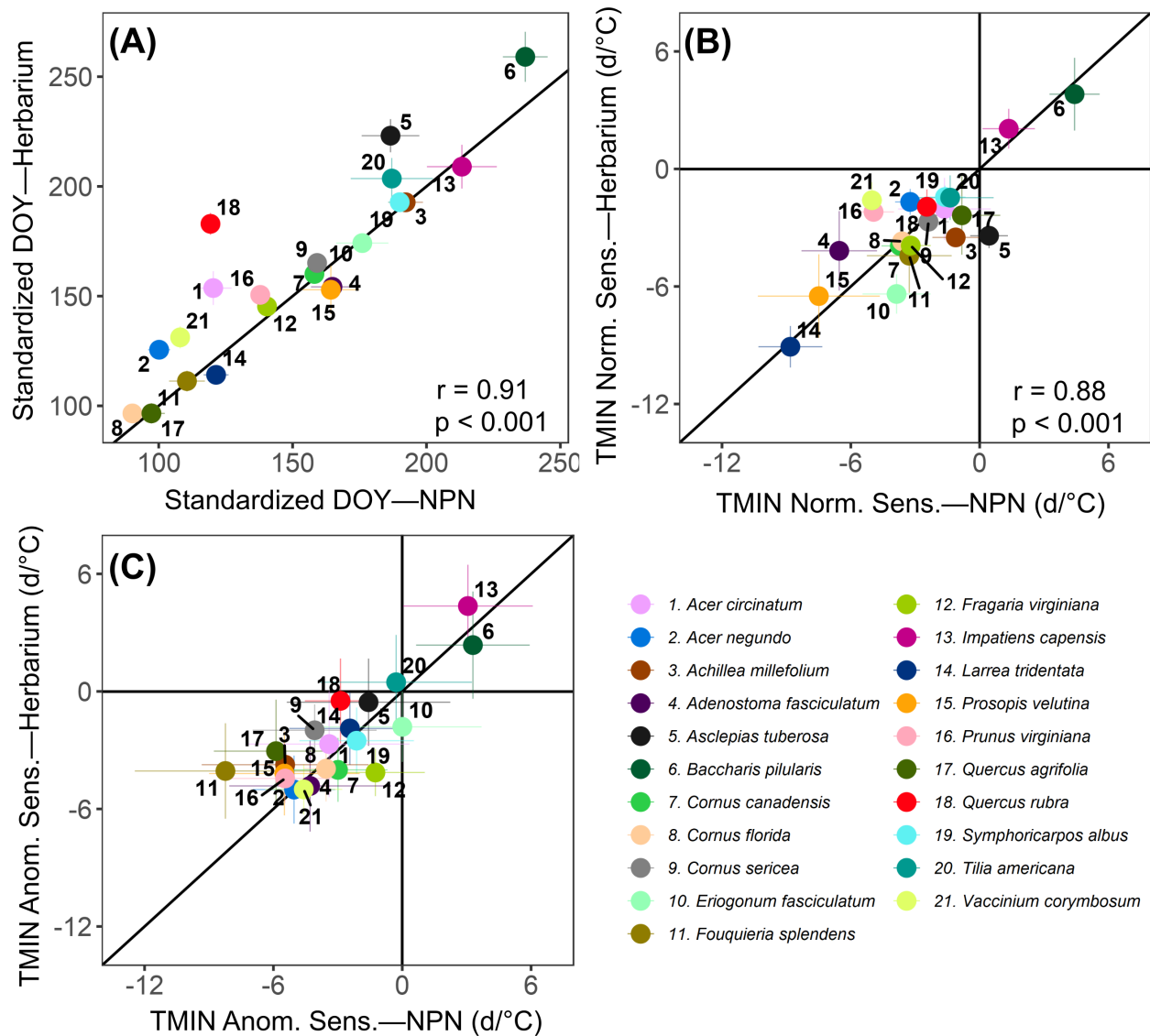


Figure 3. Comparison of species-specific estimates derived from herbarium versus USA National Phenology Network (USA-NPN) field observations for standardized flowering dates (A), flowering sensitivity to mean minimum temperature (TMIN) normal (B) and flowering sensitivity to TMIN anomaly (C). Vertical and horizontal lines around each point correspond to the standard deviation of each species-specific parameter and each dataset, and were obtained from the posterior distribution of each coefficient. Reported correlation coefficients and p-values were obtained using the minimum between the number of unique locations for each species in either the herbarium or the NPN dataset as weights. DOY = day of year of collection.

by differing climate conditions captured by each dataset. For each species, we calculated the absolute difference between herbarium- and field-based estimates of flowering time and TMIN sensitivities, and the absolute differences in mean TMIN normal and mean TMIN anomaly for both datasets. Finally, we calculated weighted Pearson correlations between absolute differences in parameter estimates and absolute differences in TMIN normal and TMIN anomaly between datasets, using the number of unique sites in the NPN or the herbarium dataset for each species as weights. All p-values were adjusted for multiple hypothesis testing using a Holm–Bonferroni correction.

Results

We found strong correlations between herbarium- and field-derived estimates for all phenological parameters. Standardized flowering times ranged from mid-Spring (early April, *Cornus florida*) to late Summer (mid-September, *Baccharis pilularis*), with very high correlation between herbarium- and field-derived estimates ($r = 0.91$, $p < 0.001$; Fig. 3A). Absolute differences between herbarium- and field-based standardized flowering dates ranged from 0 days for *Fouquieria splendens* to 64 days for *Quercus rubra* (Table 1). Overall, estimates from both datasets differed by a mean of

14 days among species, with herbarium specimens generating estimates that were, on average, 11 days later than NPN-derived estimates across species.

Estimates of sensitivity to TMIN normals and to TMIN anomalies were consistent between data types. Field- and herbarium-based estimates of sensitivity to TMIN normal were highly correlated ($r=0.88$, $p < 0.001$) and largely co-varied along a one-to-one line, indicating agreement in the magnitude of species-specific sensitivities (Fig. 3B). TMIN normal sensitivities agreed in direction (i.e. the sign of the slope coefficient; Eq. 2) for 20 out of 21 species (95%), and the only species showing discrepancies between data types (*Asclepias tuberosa*) showed a non-significant estimate of TMIN normal sensitivity for field observations (Supporting information). On average, estimates of TMIN normal sensitivity differed by $1.5 \text{ d } ^\circ\text{C}^{-1}$ among species between data types, with absolute differences ranging from $0.1 \text{ d } ^\circ\text{C}^{-1}$ for *Cornus florida* and *Tilia americana* to $4.5 \text{ d } ^\circ\text{C}^{-1}$ for *Asclepias tuberosa* (Table 1). Collectively, herbarium-based estimates were an average of $0.1 \text{ d } ^\circ\text{C}^{-1}$ more negative than field-based estimates.

Similarly, sensitivities to TMIN anomalies were significantly correlated between data types ($r=0.82$, $p < 0.001$) and tended to agree in both direction and magnitude (Fig. 3C). Sensitivities to TMIN anomalies agreed in direction for 19 out of 21 species (90%), and the two species with mismatches in direction between data types (*Eriogonum fasciculatum* and *Tilia americana*) had non-significant estimates that were very close to 0 for both data types (Fig. 3C, Supporting information). Herbarium- and field-based estimates of sensitivity to TMIN anomaly differed by an average of $1.3 \text{ d } ^\circ\text{C}^{-1}$ among species, with absolute differences ranging from $0.0 \text{ d } ^\circ\text{C}^{-1}$ for *Acer negundo* to $3.9 \text{ d } ^\circ\text{C}^{-1}$ for *Fouquieria splendens* (Table 1). Herbarium-based estimates of sensitivity to TMIN anomaly were, on average, $0.5 \text{ d } ^\circ\text{C}^{-1}$ more positive than field-based estimates.

Among species, absolute differences in mean TMIN normal and in TMIN anomaly between datasets were not significantly correlated to mismatches between herbarium-versus field-derived phenological estimates (Fig. 4). While we detected a marginally significant negative relationship between mismatches in standardized flowering dates and differences in TMIN normal between datasets, which would nonsensically indicate higher agreement between herbarium-versus field-based estimates for species showing greater differences in TMIN normal (Fig. 4A, C), such relationship was driven by a single outlier, *Quercus rubra*, exhibiting the greatest mismatch in estimated flowering time between datasets (64 d), and one of the lowest absolute differences in TMIN normal among species (0.1°C). Excluding *Q. rubra* yielded a non-significant relationship instead ($p=0.50$).

Discussion

We found strong correlations between herbarium- and field-based estimates of phenological sensitivity to mean minimum temperatures (TMIN) in both space and time for 21

species across the United States, providing the broadest demonstration to date of the reliability of herbarium specimens for estimating and comparing phenology–climate relationships across multiple species. Our results underscore the enormous promise of herbarium collections in expanding the taxonomic, geographic and temporal scope of research on phenology–climate relationships.

Our results agree with Davis et al. (2015), which found that, among 20 species in Middlesex County (Massachusetts, USA), herbarium- and field-based estimates of flowering time sensitivity to temperature agreed in direction and that mean sensitivities pooled among species did not show statistically significant differences between data types. However, Davis et al. (2015) found substantial mismatches between data types in the magnitude of sensitivity estimates (i.e. estimates were uncorrelated), resulting in disparate patterns of variation in phenological sensitivity among species for both data types. The authors attributed this result to differences in the phenological events likely captured by field observations and herbarium specimens (first flowering dates versus peak flowering dates), which can differ in their sensitivity to climate (CaraDonna et al. 2014). This interpretation is consistent with Robbirt et al. (2011), which found close agreement between field- and herbarium-derived estimates of sensitivity for a European orchid species *Ophrys sphegodes* using a large field dataset of peak (not first) flowering dates and herbarium specimens. Likewise, we found strong correlations between herbarium- and field-based estimates of TMIN sensitivity (Fig. 3), likely because NPN data allowed us to estimate median flowering dates that presumably approximate peak flowering dates, aligning better with the flowering stages captured by herbarium records.

Our study extends these results by demonstrating strong quantitative agreement between field- and herbarium-based estimates of sensitivity to climate over both space and time and across multiple species. Moreover, while Robbirt et al. (2011) focused on a single species and Davis et al. (2015) analyzed only herbaceous species with ephemeral spring and summer flowering in New England, our 21 focal species spanned a wide diversity of growth forms, life histories and native biomes, suggesting that herbarium-based estimates of phenology–climate relationships may be reliable across a wide spectrum of ecological contexts.

Despite a strong correlation, herbarium specimens produced later estimates of standardized flowering dates than did NPN observations. For most species, herbarium specimens encompassed colder TMIN conditions than NPN observations (Fig. 2). However, while this difference would predict later flowering dates in the cooler herbarium dataset (Fig. 2), differences in TMIN normal and anomaly did not explain mismatches between datasets (Fig. 4). Herbarium specimens are predominantly collected opportunistically or during sporadic botanical expeditions, which might make them more likely to represent median rather than early or late flowering individuals within a population. In turn, NPN records are assembled from regular visits to sites or individuals, which may result in capture of early flowering plants for herbaceous

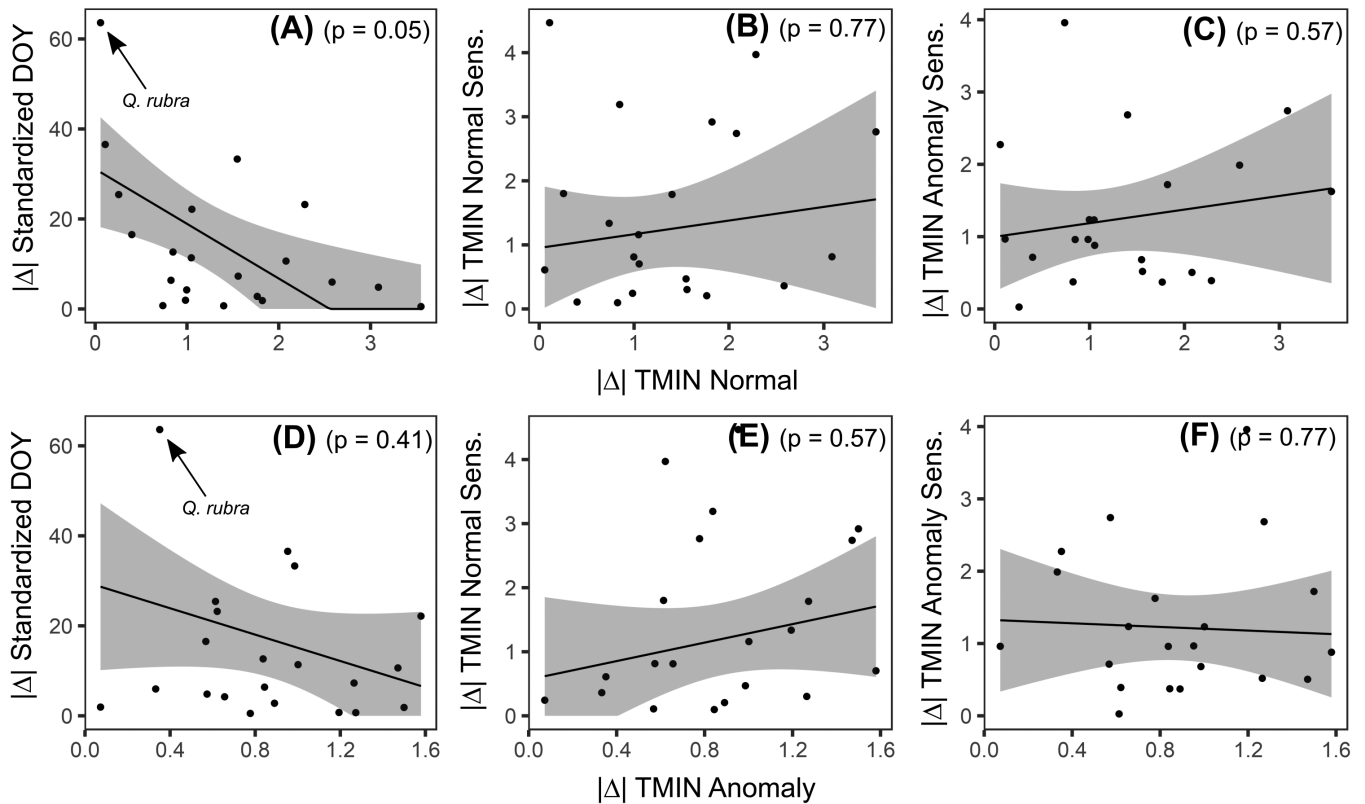


Figure 4. Correlation among 21 species in the continental United States between absolute differences ($|\Delta|$) between herbarium- versus field-derived estimates of standardized day of year of collection (DOY), of sensitivity to mean minimum temperature (TMIN) normals and of sensitivity to TMIN anomalies and absolute differences in mean TMIN normal (A–C) and TMIN anomaly (D–F). *Quercus rubra* is labeled in panels showing absolute differences in mean DOY, for which it was a clear outlier. All p values displayed were adjusted for multiple hypothesis testing using a Holm–Bonferroni correction.

species for which monitoring the same individuals across years may not be possible, and to the extent that observers might choose to monitor large and healthy trees or shrubs (which may flower early), this could be the case for woody taxa as well. Nevertheless, estimates from both datasets showed modest differences and high correlation despite marked differences in collection periods and climatic conditions.

Similarly, herbarium- and field-based estimates of sensitivities to spatiotemporal TMIN variation overwhelmingly agreed in direction and magnitude despite differences in TMIN conditions between datasets (Fig. 2B, C, 3B, C). While recent studies show that species can exhibit variation in phenological sensitivity among areas characterized by different long-term climatic conditions (Song et al. 2020, Love and Mazer 2021, Pearson et al. 2021), our results suggest that such intraspecific differences might not be substantial enough to mask patterns of among-species variation in sensitivity to TMIN in this case. Similarly, plastic phenological responses to interannual climate variation can vary intraspecifically between cool and warm periods due to non-linearities in the underlying phenology–temperature relationship (Fu et al. 2015, Güsewell et al. 2017). However, the lack of associations between mismatches in TMIN conditions and in TMIN sensitivity suggests that phenology–temperature

relationships among our focal species might be stable within the range of interannual variation encompassed here.

While we lacked enough taxa to test this statistically, we could not discern any relationships between species-level characteristics and the degree of mismatch between herbarium versus field estimates. For example, while the species that showed the greatest mismatches for different phenological parameters consisted of a mix of evergreen and deciduous woody species (and a few herbs) from various western and eastern ecoregions, so did groups of species showing the smallest mismatches (Table 1). Likewise, we did not discern clear taxonomic patterns in mismatches between datasets. For example, while species in the genera *Cornus* and *Quercus*, respectively, showed some of the smallest mismatches in TMIN normal sensitivity and the greatest mismatches in TMIN anomaly sensitivity, congeners *Cornus florida* and *C. sericea* were respectively among the species showing the smallest and greatest mismatches in TMIN anomaly, obfuscating whether the reliability of herbarium-derived estimates may vary taxonomically.

NPN observations and herbarium collections might exhibit similar biases not examined in this study. For example, specimens might be collected and NPN observations conducted at easily accessible sites near roads or at low elevations that may inaccurately represent the overall

environmental conditions and phenology observed throughout a species' range (Daru et al. 2018, Meineke and Daru 2021). Additionally, we detected large differences in sensitivity estimates for some species and substantial uncertainty in parameter estimation (especially for sensitivity to TMIN anomaly; Fig. 3C, Supporting information), suggesting that herbarium-derived sensitivities for some species may lead to different conclusions from field observations or require much greater sample sizes than employed here for accurate estimation. Nevertheless, within the geographic and climatic space and the ecological diversity sampled in this study, our results demonstrate that herbarium specimens can uncover patterns of variation in phenology–climate relationships largely equivalent to those generated using field observations, suggesting that herbarium-based estimates may be generally robust to potential error or bias in specimen collection dates as proxies of peak flowering time.

Future directions

We provide strong evidence of the reliability of herbarium specimens as resources with which to study phenological responses to spatiotemporal climate variation among species. However, our study was constrained by the availability of well-represented species in the NPN and herbarium datasets, preventing statistical comparison of the reliability of herbarium-based estimates among, for example, species with different life history traits. Future studies could leverage the growing number of digitized collections across the United States to identify additional species that are well represented in the NPN or other observational datasets and that might facilitate such analyses. Additionally, our study focused on a single component of the flowering phenology of a species (peak flowering); further research could determine whether specimens can generate reliable estimates of sensitivity for flowering onset or termination (which can show differing responses to climate; CaraDonna et al. 2014), or for different life-cycle stages altogether. Phenological data from herbarium specimens are usually limited to presence–absence of reproductive structures, providing coarse information on the reproductive stage of specimens. Ongoing efforts to automate scoring of reproductive structures in herbarium sheets (Pearson et al. 2020) combined with new metrics that provide fine-grained information of the reproductive status of herbarium specimens (Love et al. 2019, Goëau et al. 2020) might eventually enable sensitivity analyses for a wide range of phenological events and stages.

Acknowledgements – We thank the many herbaria, including staff and volunteers, who collected, curated and digitized the vast volumes of herbarium specimens leveraged for this study (see Supporting information for full listing). We thank Devin Gamble, Cameron Hannah-Bick, Lisa Kim, Helen Payne and Zoe Woods for their constructive feedback on previous iterations of this manuscript. T.R.P. thanks the University of California, Santa

Barbara, for fellowship support in the year during which this paper was completed.

Funding – This work was supported by the National Science Foundation, grant no. DEB-1556768 (to S.J.M. and I.W.P.).

Author contributions

Tadeo H. Ramirez-Parada: Conceptualization (equal); Data curation (equal); Formal analysis (lead); Methodology (lead); Visualization (lead); Writing – original draft (lead); Writing – review and editing (lead). **Isaac W. Park:** Conceptualization (supporting); Data curation (equal); Funding acquisition (lead); Methodology (supporting); Writing – original draft (supporting); Writing – review and editing (supporting). **Susan J. Mazer:** Conceptualization (equal); Funding acquisition (lead); Methodology (supporting); Project administration (lead); Writing – original draft (supporting); Writing – review and editing (supporting).

Transparent peer review

The peer review history for this article is available at <<https://publons.com/publon/10.1111/ecog.06173>>.

Data availability statement

Data are available from the Dryad Digital Repository: <<https://doi.org/10.25349/D9TK64>> (Ramirez-Parada et al. 2022).

Supporting information

The Supporting information associated with this article is available with the online version.

References

- Anderson, J. T. et al. 2012. Phenotypic plasticity and adaptive evolution contribute to advancing flowering phenology in response to climate change. – *Proc. R. Soc. B* 279: 3843–3852.
- Calinger, K. M. et al. 2013. Herbarium specimens reveal the footprint of climate change on flowering trends across north-central North America. – *Ecol. Lett.* 16: 1037–1044.
- CaraDonna, P. J. et al. 2014. Shifts in flowering phenology reshape a subalpine plant community. – *Proc. Natl Acad. Sci. USA* 111: 4916–4921.
- Chuine, I. 2010. Why does phenology drive species distribution? – *Phil. Trans. R. Soc. B* 365: 3149–3160.
- Cleland, E. E. et al. 2012. Phenological tracking enables positive species responses to climate change. – *Ecology* 93: 1765–1771.
- Cook, B. I. et al. 2012. Sensitivity of spring phenology to warming across temporal and spatial climate gradients in two independent databases. – *Ecosystems* 15: 1283–1294.
- Daru, B. H. et al. 2018. Widespread sampling biases in herbaria revealed from large-scale digitization. – *New Phytol.* 217: 939–955.
- Davis, C. C. et al. 2015. Herbarium records are reliable sources of phenological change driven by climate and provide novel insights into species' phenological cueing mechanisms. – *Am. J. Bot.* 102: 1599–1609.

- Delgado, M. del M. et al. 2020. Differences in spatial versus temporal reaction norms for spring and autumn phenological events. – *Proc. Natl Acad. Sci. USA* 117: 31249–31258.
- Fu, Y. H. et al. 2015. Declining global warming effects on the phenology of spring leaf unfolding. – *Nature* 526: 104–107.
- Goëau, H. et al. 2020. A new fine-grained method for automated visual analysis of herbarium specimens: a case study for phenological data extraction. – *Appl. Plant Sci.* 8: e11368.
- Güsewell, S. et al. 2017. Changes in temperature sensitivity of spring phenology with recent climate warming in Switzerland are related to shifts of the pre-season. – *Global Change Biol.* 23: 5189–5202.
- Heberling, J. M. et al. 2019. The changing uses of herbarium data in an era of global change: an overview using automated content analysis. – *BioScience* 69: 812–822.
- Jones, C. A. and Daehler, C. C. 2018. Herbarium specimens can reveal impacts of climate change on plant phenology: a review of methods and applications. – *PeerJ* 6: e4576.
- Love, N. L. R. and Mazer, S. J. 2021. Region-specific phenological sensitivities and rates of climate warming generate divergent temporal shifts in flowering date across a species' range. – *Am. J. Bot.* 108: 1873–1888.
- Love, N. L. R. et al. 2019. A new phenological metric for use in pheno-climatic models: a case study using herbarium specimens of *Streptanthus tortuosus*. – *Appl. Plant Sci.* 7: e11276.
- Meineke, E. K. et al. 2018. The unrealized potential of herbaria for global change biology. – *Ecol. Monogr.* 88: 505–525.
- Menzel, A. et al. 2020. Climate change fingerprints in recent European plant phenology. – *Global Change Biol.* 26: 2599–2612.
- Meineke, E. K. and Daru, B. H. 2021. Bias assessments to expand research harnessing biological collections. – *Trends Ecol. Evol.* 36: 1071–1082.
- Miller-Rushing, A. J. and Primack, R. B. 2008. Global warming and flowering times in Thoreau's concord: a community perspective. – *Ecology* 89: 332–341.
- Miller-Rushing, A. J. et al. 2006. Photographs and herbarium specimens as tools to document phenological changes in response to global warming. – *Am. J. Bot.* 93: 1667–1674.
- Munson, S. M. and Long, A. L. 2017. Climate drives shifts in grass reproductive phenology across the western USA. – *New Phytol.* 213: 1945–1955.
- Omernik, J. M. and Griffith, G. E. 2014. Ecoregions of the conterminous United States: evolution of a hierarchical spatial framework. – *Environ. Manage.* 54: 1249–1266.
- Panchen, Z. A. et al. 2019. Patterns and biases in an Arctic herbarium specimen collection: implications for phenological research. – *Appl. Plant Sci.* 7: e01229.
- Park, I. W. 2012. Digital herbarium archives as a spatially extensive, taxonomically discriminate phenological record; a comparison to MODIS satellite imagery. – *Int. J. Biometeorol.* 56: 1179–1182.
- Park, I. W. 2014. Impacts of differing community composition on flowering phenology throughout warm temperate, cool temperate and xeric environments. – *Global Ecol. Biogeogr.* 23: 789–801.
- Park, I. W. and Mazer, S. J. 2018. Overlooked climate parameters best predict flowering onset: assessing phenological models using the elastic net. – *Global Change Biol.* 24: 5972–5984.
- Pearson, K. D. et al. 2021. Phenological trends in the California poppy *Eschscholzia californica*: digitized herbarium specimens reveal intraspecific variation in the sensitivity of flowering date to climate change. – *Madroño* 68: 343–359.
- Pearson, K. D. et al. 2020. Machine learning using digitized herbarium specimens to advance phenological research. – *BioScience* 70: 610–620.
- Primack, D. et al. 2004. Herbarium specimens demonstrate earlier flowering times in response to warming in Boston. – *Am. J. Bot.* 91: 1260–1264.
- Ramirez-Parada, T. H. et al. 2022. Data from: Herbarium specimens provide reliable estimates of phenological responses to climate at unparalleled taxonomic and spatiotemporal scales. – *Dryad Digital Repository*, <<https://doi.org/10.25349/D9TK64>>.
- Reeves, P. H. and Coupland, G. 2000. Response of plant development to environment: control of flowering by daylength and temperature. – *Curr. Opin. Plant Biol.* 3: 37–42.
- Renner, S. S. and Zohner, C. M. 2018. Climate change and phenological mismatch in trophic interactions among plants, insects and vertebrates. – *Annu. Rev. Ecol. Evol. Syst.* 49: 165–182.
- Robbirt, K. M. et al. 2011. Validation of biological collections as a source of phenological data for use in climate change studies: a case study with the orchid *Ophrys sphegodes*. – *J. Ecol.* 99: 235–241.
- Schwartz, M. D. et al. 2012. From Caprio's lilacs to the USA National Phenology Network. – *Front. Ecol. Environ.* 10: 324–327.
- Song, Z. et al. 2020. Flowering phenology of a widespread perennial herb shows contrasting responses to global warming between humid and non-humid regions. – *Funct. Ecol.* 34: 1870–1881.
- Spellman, K. V. and Mulder, C. P. H. 2016. Validating herbarium-based phenology models using citizen-science data. – *BioScience* 66: 897–906.
- Stan Development Team 2020. RStan: the R interface to Stan. – R package ver. 2.21.2, <<http://mc-stan.org/>>.
- Tang, J. et al. 2016. Emerging opportunities and challenges in phenology: a review. – *Ecosphere* 7: e01436.
- Templ, B. et al. 2018. Pan European phenological database (PEP725): a single point of access for European data. – *Int. J. Biometeorol.* 62: 1109–1113.
- Wang, T. et al. 2016. Locally downscaled and spatially customizable climate data for historical and future periods for North America. – *PLoS One* 11: e0156720.
- Wolkovich, E. M. et al. 2014. Progress towards an interdisciplinary science of plant phenology: building predictions across space, time and species diversity. – *New Phytol.* 201: 1156–1162.
- Willis, C. G. et al. 2017. Old plants, new tricks: phenological research using herbarium specimens. – *Trends Ecol. Evol.* 32: 531–546.