



Improving phenology predictions for sparsely observed species through fusion of botanical collections and citizen-science

Lucien Fitzpatrick^{a,*}, Perry J. Giambuzzi^a, Alena Spreitzer^b, Brendon Reidy^a, Shannon M. Still^c, Christine R. Rollinson^a

^a Center for Tree Science, The Morton Arboretum, Lisle, IL 60532, USA

^b The University of Chicago, Chicago, IL 60637, USA

^c UC Davis Arboretum & Public Garden, Davis, CA 95616, USA

ARTICLE INFO

Keywords:

Ecological modeling
Spatial variability
Bayesian modeling
Climate change
Rare species
Vulnerability

ABSTRACT

Describing patterns of plant phenology through models has been critical for quantifying species responses to climate change and forecasting future vegetation impacts. However, many species remain unincluded in large analyses because they are poorly represented in the large public or citizen science datasets that form the foundation of these efforts. Botanical living collections are often key resources that facilitate study of rare and sparsely observed species, but alone are insufficient to predict species phenology throughout their observed ranges. We investigate whether predictions for rare and data-poor species observed at a single site can be improved by leveraging observations of similar taxa observed at multiple locations. We combined observations of oak (*Quercus*) budburst and leaf out from one botanical garden with a subset of congeneric species observed in the USA-NPN citizen science dataset using Bayesian hierarchical modeling. We show that including USA-NPN observations into a simple thermal time model of budburst and leaf out did not reduce geographic bias in model predictions over models parameterized only with single-site observations. However, using USA-NPN data to add non-taxonomic spatial covariates to the thermal time model improved model performance for all species, including those only observed at a single site. Living collections at botanical gardens provide valuable opportunities to observe rare or understudied species, but are limited in geographic scope. National-scale citizen science observations that capture the spatial variability of related or ecologically similar taxa can be combined with living collections data to improve predictions of species of conservation concern across their native range.

1. Introduction

Plant phenological responses to spring temperatures and warming trends over recent decades is one of the clearest indicators of climate change impacts on plants and ecosystems across the globe [1,2]. As many as 78% of European cold-deciduous plant species are now showing earlier budburst, leaf out, or flowering timing than 50 years ago [2]. The mechanistic ties among plant phenology, species reproductive success, and individual survival have allowed for improved understanding of climate vulnerability for many temperate tree species [3,4]. However, phenology research and climate vulnerability assessments are typically only possible for common temperate deciduous species in the northern hemisphere where there is high data availability from long-term research, experiments, and citizen science studies [5,6]. The climate risk for species with narrow ranges, common in rare and threatened taxa, is often more difficult to assess [7,8] and different methods of evaluating this vulnerability may provide diverging results [9]. Providing

additional information to help manage rare or restricted-distribution plants is of interest to the botanical garden and conservation communities as these taxa are generally considered to have a higher risk of being impacted by anthropogenic events [10], are often phylogenetically diverse [10], and the identification of threats of extirpation to populations of rare species can be helpful in informing conservation priorities [10–12].

Data limitations describing biological responses to climate and the environment is a common and pervasive challenge for robust threat assessments of rare species. Incorporating the effects of climate change vulnerability into established threat programs such as the IUCN Red List is a particular challenge for many species due to a lack of available data on the effects of climate variability and change on individual and population-scale biology [13]. Although phenological observations are relatively easy to make, most taxa of current conservation concern are not incorporated into large data networks due to challenges of access and identification that complicate data reliability [14]. For

* Corresponding author.

E-mail address: lfitzpatrick@mortonarb.org (L. Fitzpatrick).

<https://doi.org/10.1016/j.ecochg.2021.100032>

Received 23 December 2020; Received in revised form 17 September 2021; Accepted 24 September 2021

Available online 4 October 2021

2666-9005/© 2021 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

example, The USA National Phenology Network (USA-NPN) Nature's Notebook citizen science program is one of the largest phenological monitoring programs and has been used to make species-level forecasts of activity for dozens of both plant and animal species to be used for both research and management-focused audiences [15–17]. However, of the 91 *Quercus* L. species in the US, only 29 are included in the USA-NPN observing program, and all except two (*Q. engelmannii* Greene, *Q. lobata* Née) are considered species of “least concern” in the Red List of US Oaks [18]. Existing information on phenology and climate responses for rare and endemic species is often the product of isolated sampling efforts, inconsistent methodologies, and environmental conditions that challenge species comparisons and extrapolation to other conditions [13,19,20]. To address these limitations, analytical methods that combine datasets, particularly those that borrow strength from data-rich related or functionally similar species over broader regions, may enhance our ability to predict phenology for data-poor species.

Botanical garden living collections and herbaria are able to help fill critical data needs for rare and restricted-distribution species that are logistically challenging to observe *in situ* and are typically absent from large-scale and citizen science datasets such as the USA-NPN [21–23]. Botanical gardens and arboreta have long been key sources of taxonomic and phylogenetic comparisons of traits like phenology across the diversity of taxa grown in common garden environments and have been essential for global biodiversity work [24,25]. Through just two years of observations, data from botanical gardens have highlighted how phylogenetic and trait patterns explaining spring leaf out do not hold for autumn senescence [25,26]. However, the ability of gardens to address information needs for individual species responses to both climate variability and change has been difficult due to the common limitations of single-site studies to describe the full geographic or climatic range of *in situ* individuals making the extrapolation of the *ex situ* conditions tenuous [27]. Bolstering the comparably-narrow climate and geography of botanical collections with the large geographic and environmental extent of citizen science phenology programs has the potential to improve the robustness of phenological predictions for rare species in habitats where direct observation can be infeasible.

Phenological models that incorporate species-level biological threshold responses to weather and climate conditions have been used to accurately forecast phenology in near real time [16,17] but these studies and applications typically focus on common data-rich taxa. Current phenological models range in complexity from thermal time models that use a single parameter, such as a growing degree-day threshold, to more complex multivariate models that account for additional cues such as chilling and spatio-temporal variability in light and winter temperatures [27–29]. Multivariate models often out-perform simpler models when predicting phenological events over large areas or outside of the data collection site [28,30,31], but because the data needed to parameterize models scales non-linearly with complexity, accurate phenology modeling and prediction is typically not possible for many rare or uncommon species when analyzed on their own [31,32]. Historically, hierarchical mixed models have been widely used to borrow statistical strength across unbalanced datasets and improve predictions in many applications, but they have not yet been widely applied to facilitate the use of large citizen science datasets to improve modeling for rare and data-limited species [13,14].

In this study, we test whether phenological models for rare and data-poor species observed at a single site can be improved by leveraging observations of phenologically similar, congeneric species from a large, distributed citizen science dataset. We modeled leaf out and budburst of seven temperate deciduous oak (*Quercus*) species using a hierarchical Bayesian thermal time model with and without spatial covariates to combine observations from The Morton Arboretum (TMA) and the USA National Phenology Network (USA-NPN). Three species were sparsely-observed in USA-NPN (<25 sites) and were modeled using data from only TMA (“TMA Only” group) and validated using all available USA-NPN observations to test the ability of leveraging data from

congeneric species to improve phenology models: *Q. imbricaria* Michx., *Q. montana* Wiild.(syn *Q. prinus* L. *p.p.*, *nom. utique rej.*), and *Q. velutina* Lam. (Table 1). To test the power of using large, spatially-distributed datasets for improving predictions far outside of the available training data, USA-NPN data for the remaining four congeneric species was included in models to explain phenological patterns over larger spatial scales (“TMA + NPN” group): *Q. alba* L., *Q. macrocarpa* Michx., *Q. palustris* Münchh., *Q. rubra* L.. Predicted phenology for species in the TMA + NPN group were evaluated using a random 25% of USA-NPN sites excluded from model fitting (Table 1). We modeled leaf out and budburst phenology for all seven species using four hierarchical phenological models: 1) TMA data only (TMA); 2) TMA and USA-NPN data (COMB); 3) TMA and USA-NPN data and a latitude spatial covariate (LAT); 4) TMA and USA-NPN data and a winter temperature spatial covariate (WT). All models and species were evaluated based on root mean square error (RMSE) and spatial bias in prediction accuracy using USA-NPN observations withheld from model fitting.

2. Method

2.1. Data sets descriptions and data cleaning

We analyzed phenology of seven *Quercus* species native to the Eastern U.S. observed both at The Morton Arboretum's Oak Collection (TMA) and at least five other USA-NPN sites (Fig. 1, Supplemental Table S1 & S2). Of the 16 species common to both TMA and USA-NPN, seven species met our criteria for inclusion in our study following data cleaning, described below. TMA data were collected approximately weekly in 2018 and 2019 following USA-NPN protocols [33]. Additional species are observed at TMA, but are not included in this study due to lack of available independent observations for prediction evaluation. For USA-NPN data, leaf out and budburst individual phenometrics [34] for 2008–2019 were downloaded from Nature's Notebook using the *rnpn* package in R [35]. Individual phenometrics are calculated by USA-NPN and indicate onset and end date for phenophases such as budburst and leaf out [34]. USA-NPN sites are distributed across the United States and are often concentrated in population dense areas (Fig. 1).

Both TMA and USA-NPN datasets were cleaned prior to analysis following approaches used in other studies [36] and TMA observations were removed from the USA-NPN dataset (site 26202) to avoid observation duplication. We restricted our analyses to exclude observations of budburst or leaf out presence that had not been preceded by an absence observation within the prior 10 days, which removed 39% of the initial 8243 individual budburst and leaf out USA-NPN phenometrics for our study taxa. For individuals with multiple budburst or leaf out entries per year, we used the first instance. We also removed observations after the summer solstice (day of year 172) to restrict observations to start-of-season phenophase activity. Finally, we excluded outliers in the USA-NPN dataset by removing any observations outside of three standard deviations from the mean day of year for each phenophase for each species. Final number of unique budburst and leaf out observations, trees, and sites for both TMA and USA-NPN can be found in Tables S1 and S2.

The objective of this study was to evaluate the ability of observations from related taxa in geographically dispersed datasets such as USA-NPN to improve predictions for rare or sparsely-observed species from a single site. To achieve this objective, we categorized species used in this study into two groups. Species in the “TMA + NPN” group used observations from both TMA and USA-NPN to parameterize species-level effects in all models (Section 3), while the “TMA Only” group withheld all USA-NPN observations for model validation (Fig. 1). Thus, the “TMA Only” group provides a means of evaluating the likely prediction error and spatial bias for other species observed at sites such as TMA, but for which no independent prediction evaluation is currently possible. Because our goal was to focus on the ability to improve predictions for rare and endemic species, we chose species observed in less than 20 sites for the “TMA Only” group: *Q. velutina*,

Table 1

Description of how data from living collections at The Morton Arboretum's (TMA) and The USA National Phenology Network (USA-NPN) are used in multi-site models of leaf out and budburst in our study. For species with observations at more than 20 sites (*Q. alba*, *Q. macrocarpa*, *Q. palustris*, *Q. rubra*; collectively termed "TMA + NPN"), 25% of USA-NPN sites for each species were withheld from model fitting for model validation while the remaining sites used TMA data in the model fitting process (calibration). All USA-NPN sites for the remaining three species were withheld from model fitting to test the ability of using congeneric observations to improve predictions made with species observations from a single site ("TMA Only"; *Q. imbricaria*, *Q. montana*, *Q. velutina*).

Species	Data for model parameterization	Data for Validation
<i>Q. alba</i>	TMA + 75% USA-NPN sites	Random 25% sample of USA-NPN sites
<i>Q. macrocarpa</i>	TMA + 75% USA-NPN sites	Random 25% sample of USA-NPN sites
<i>Q. palustris</i>	TMA + 75% USA-NPN sites	Random 25% sample of USA-NPN sites
<i>Q. rubra</i>	TMA + 75% USA-NPN sites	Random 25% sample of USA-NPN sites
<i>Q. imbricaria</i>	TMA	All USA-NPN sites
<i>Q. montana</i>	TMA	All USA-NPN sites
<i>Q. velutina</i>	TMA	All USA-NPN sites

Q. montana, *Q. imbricaria* (Table S1, S2). To provide context for the performance of phenology predictions made from the single-site TMA group, we randomly excluded one-fourth of USA-NPN sites for each remaining "TMA + NPN" species (*Q. alba*, *Q. macrocarpa*, *Q. palustris*, *Q. rubra*; Tables S1, S2) for model validation and used the remaining observations in the model calibration and fitting process (Fig. 1). One additional species, *Q. ilicifolia* Wangenh., met the criteria for inclusion in our study as a "TMA Only" validation species, but the species parameters were unable to converge in our models (described below).

2.2. Meteorological Data

Weather data used in this study was acquired for TMA and all USA-NPN sites from Daymet using the *daymetr* package [38]. Daymet is a national weather collection service that provides gridded estimates of daily weather parameters at 1 km spatial resolution [39,40]. We used daily minimum and maximum temperatures to calculate the daily mean temperature value which was then used to calculate accumulated growing degree days (GDD) and winter season mean temperature (WT). GDD was calculated by subtracting a base temperature of 5 °C from the daily mean temperature with any negative values being set to 0. The GDD value associated with a budburst or leaf out event was the sum of GDD between January 1st and the date of that phenological event observation (e.g. day of first budburst or leaf out). Winter season mean temperature (WT) was calculated as the mean daily temperature from Jan. 1st through March 1st. Winter season mean temperature at TMA was -3.97 °C in 2018 and -5.19 °C in 2019. WT of associated with our USA-NPN observations had a median (range) of -0.6 °C (-16.3 °C - 13.3 °C).

3. Calculation

We used a series of four models to evaluate the ability of leveraging additional phenological observations of related taxa from large data networks such as USA-NPN to improve predictions made from data at a single site, TMA. Two models are simple thermal-time models with species-level cumulative growing degree-day thresholds (base temperature 5 °C) that were parameterized with two different datasets: 1) data for all seven species from The Morton Arboretum (TMA model); and 2) TMA observations for all seven species plus additional observations for four species from USA-NPN (COMB model). Because the power of big data such as USA-NPN often comes from its ability to describe spatial variation in patterns rather than detail for many species [14], we create two additional models where USA-NPN observations were used to add one of two global spatial covariates: latitude (LAT model) or winter temperature (WT model). In all models, the accumulated growing degree-days at the time of first observed budburst or leaf out for each individual in each year was used as the response variable (GDD) and budburst and leaf out were modeled separately. More detail on all model structures are described below. Models were constructed using the Gibbs sampler

software Jags [41], and the *rjags* [42] and *coda* [43] packages. All models and analyses were performed using R [44] and Rstudio [45]. All code is available in the following Github repository: https://github.com/MortonArb-ForestEcology/Collections_phenology_vulnerability.

3.1. TMA thermal time model (single-site)

Both the TMA and COMB models are simple GDD thermal time models that differ in the datasets used in parameterization of those models. The TMA model (Eq. (1)) used all available observations of budburst or leaf out (GDD_{ijl}) from The Morton Arboretum for all seven study species and included fixed effects to describe species differences in GDD thresholds for budburst and leaf out (μ_{spp}) and individual precision around those effects (μ_{ind}). Model structure for the TMA was as follows:

$$\text{GDD}_{ijl} \sim \mathcal{N}(\mu_{ind(jl)}, \tau_{obs}) \quad (1)$$

where GDD_{ijl} was the accumulated GDD at the time of phenological event observation (*i*) of an individual (*j*) that belongs to species (*l*). The term μ_{obs} measured the overall precision (error) in our data and $\mu_{ind(jl)}$ described individual effects of individual (*j*) belonging to species (*l*). The term $\mu_{ind(jl)}$ was defined as a normal distribution:

$$\mu_{ind(jl)} \sim \mathcal{N}(\mu_{spp(l)}, \tau_{ind}) \quad (2)$$

where $\mu_{spp(l)}$ was the species effect for each species (*l*) and τ_{ind} was the precision of individuals. The species effect term, $\mu_{spp(l)}$, was also defined as a normal distribution:

$$\mu_{spp(l)} \sim \mathcal{N}(\alpha_1, \tau_{spp(l)}) \quad (3)$$

where α_1 was the prior for each species (*l*) and $\mu_{spp(l)}$ was the precision of each species (*l*). The fixed species effect had a uniform prior:

$$\tau_1 \sim \text{uniform}(0, 1000) \quad (4)$$

Precision parameters were given gamma priors because gamma distributions are nonnegative and are conjugate to normal distributions:

$$\tau_{obs}, \tau_{ind}, \tau_{spp(l)} \sim \text{gamma}(.01, .1) \quad (5)$$

3.2. COMB thermal time model

The COMB model (Eq. (6)) was fit using all available TMA observations for all seven species and observations from three-quarters of the available USA-NPN sites for each of the four "TMA + NPN" species (*Q. alba*, *Q. macrocarpa*, *Q. palustris*, *Q. rubra*). The COMB model had a similar structure as the TMA model, but with an additional hierarchical location effect that the single-site TMA model lacked without including a spatial covariate in the GDD model. The structure for the COMB model was as follows:

$$\text{GDD}_{ijkl} \sim \mathcal{N}(\mu_{ind(jkl)}, \tau_{obs}) \quad (6)$$

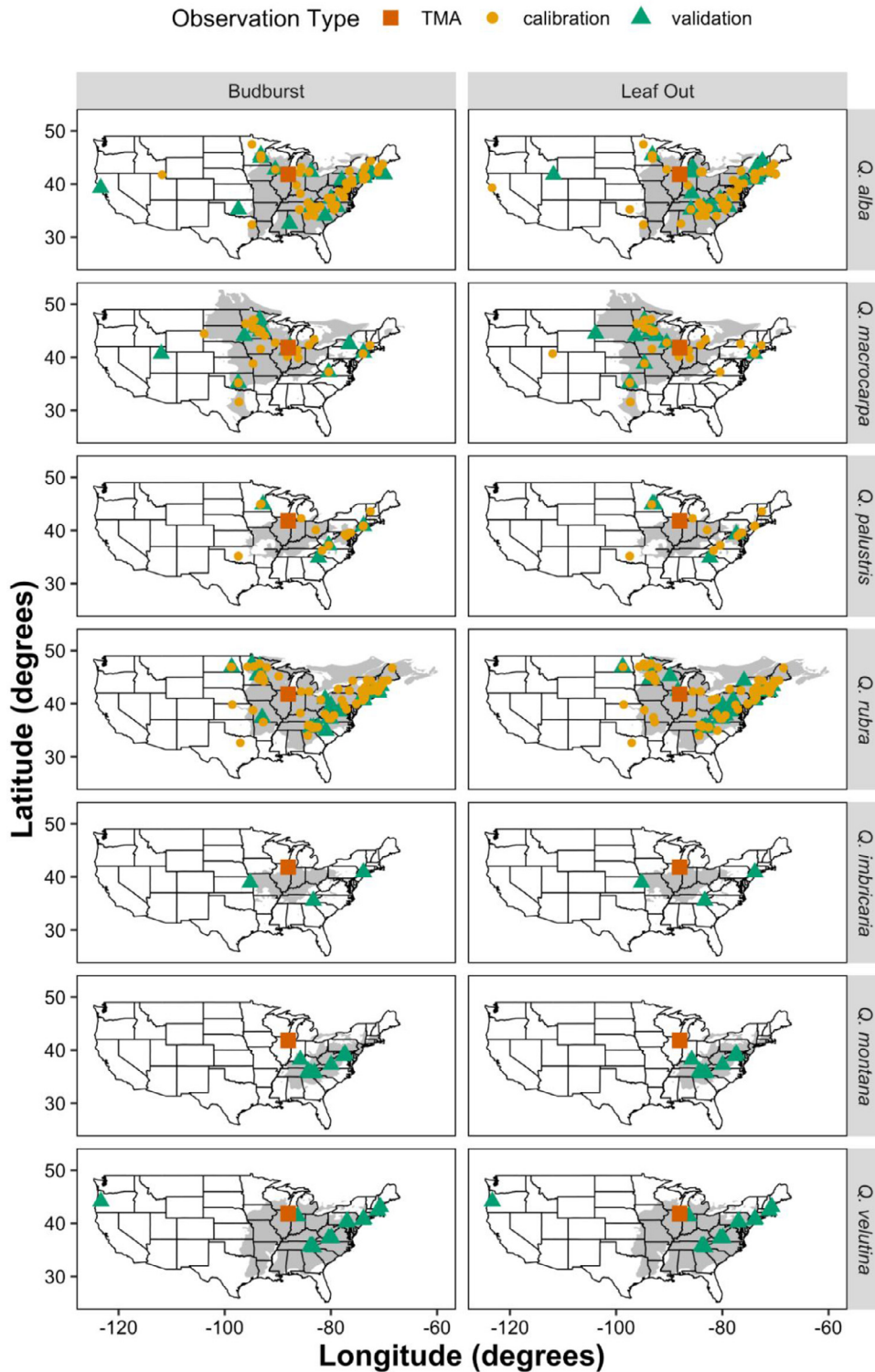


Fig. 1. Locations of all The Morton Arboretum (TMA) (orange square) and USA National Phenology Network (USA-NPN) observations for seven oak (*Quercus*) species used in this study. For species with observations at more than 20 sites (*Q. alba*, *Q. macrocarpa*, *Q. palustris*, *Q. rubra*; collectively termed “TMA + NPN”), 25% of USA-NPN sites for each species were withheld from model fitting for model validation (green triangles) while the remaining sites used TMA data in the model fitting process (calibration; yellow circles). All USA-NPN sites for the remaining three species were withheld from model fitting to test the ability of using congeneric observations to improve predictions made with species observations from a single site (“TMA Only”; *Q. imbricaria*, *Q. montana*, *Q. velutina*). Native species ranges from Little [37] are shown in gray.

where GDD_{ijkl} was the accumulated GDD at the time of phenological event observation (i) of an individual (j) observed at site (k) that belongs to species (l). The term τ_{obs} measured the overall precision (error) in the observations and $\mu_{ind(jkl)}$ described individual effects of individual (j) at site (k) belonging to species (l). The term $\mu_{ind(jkl)}$ was defined as a normal distribution:

$$\mu_{ind(jkl)} \sim N(\mu_{loc(kl)}, \tau_{ind}) \quad (7)$$

where $\mu_{loc(kl)}$ was the location effect at site (k) for each species (l) and τ_{ind} was the precision of individuals. The location effect term ($\mu_{loc(kl)}$) was defined as a normal distribution:

$$\mu_{loc(kl)} \sim N(\mu_{spp(l)}, \tau_{loc(kl)}) \quad (8)$$

where $\mu_{spp(l)}$ was the species effect for each species (l) and $\tau_{loc(kl)}$ was the precision at site (k) for each species (l). The species effect term,

$\mu_{\text{spp}(l)}$, was also defined as a normal distribution:

$$\mu_{\text{spp}(l)} \sim \mathbf{N}(\alpha_1, \tau_{\text{spp}(l)}) \quad (9)$$

where α_1 was the prior for each species (l) and $\tau_{\text{spp}(l)}$ was the precision attributed to each species (l). The fixed species effect had a uniform prior:

$$\alpha_1 \sim \text{uniform}(0, 1000) \quad (10)$$

Precision parameters were all given gamma priors because gamma distributions are nonnegative and conjugate to normal distributions:

$$\tau_{\text{obs}}, \tau_{\text{ind}}, \tau_{\text{loc}(k)}, \tau_{\text{spp}(l)} \sim \text{gamma}(.01, .1) \quad (11)$$

3.3. Spatial covariate models: LAT, WT

We also created two models that included similar effects describing species-level sources of variation in GDD thresholds (TMA model), but also included global spatial covariate effects to describe spatial variation in GDD cues across the geographic extent of available observation data, regardless of species identity (Eq. (12)). This differed from the COMB model which simply added more species-specific data because it used the USA-NPN data to capture the spatial extent of the USA-NPN data to describe spatial variation across the data set independent of species. These models used the same data inputs as the COMB model in conjunction with either a latitude or winter temperature spatial covariate taken from the location of observation. The structure of both the LAT and WT models were the same:

$$\text{GDD}_{\text{ijkl}} \sim \mathbf{N}(\mu_{\text{ind}(jl)} + (\alpha_{\text{ind}(jk)} + \beta_{\text{ind}(jk)} \times \text{SPAT}), \tau_{\text{obs}}) \quad (12)$$

where GDD_{ijkl} was the accumulated GDD at the time of phenological event observation (i) of an individual (j) observed at site (k) belonging to species (l). The term τ_{obs} described the overall precision in our data and $\mu_{\text{ind}(jl)}$ described taxonomic effects of each individual (j) belonging to species (l). The term $\alpha_{\text{ind}(jk)}$ was the spatial intercept for each individual (j) at site (k), and the term $\beta_{\text{ind}(jk)}$ was the spatial slope for each individual (j) at site (k). SPAT represented the spatial covariate that was either the latitude (LAT model) or the winter season growing mean temperature (WT model) of the observation site. The three terms $\mu_{\text{ind}(jl)}$, $\alpha_{\text{ind}(jk)}$, and $\beta_{\text{ind}(jk)}$ had unique hierarchies that were defined as follows. As in the TMA model, $\mu_{\text{ind}(jl)}$, was defined as a normal distribution:

$$\mu_{\text{ind}(jl)} \sim \mathbf{N}(\mu_{\text{spp}(l)}, \tau_{\text{ind}}) \quad (13)$$

where $\mu_{\text{spp}(l)}$ was the species effect for each species (l) and τ_{ind} was the precision of individuals. The species effect term, $\mu_{\text{spp}(l)}$, was defined as a normal distribution:

$$\mu_{\text{spp}(l)} \sim \mathbf{N}(\alpha_1, \tau_{\text{spp}(l)}) \quad (14)$$

where α_1 was the species effect prior for each species (l) and $\tau_{\text{spp}(l)}$ was the precision of each species (l). The spatial intercept, $\alpha_{\text{ind}(jk)}$, was defined as a normal distribution:

$$\alpha_{\text{ind}(jk)} \sim \mathbf{N}(\alpha_{\text{loc}(k)}, \tau_{\alpha \text{ind}}) \quad (15)$$

where $\alpha_{\text{loc}(k)}$ was the location intercept effect at each site (k) and $\tau_{\alpha \text{ind}}$ was the precision of the spatial intercept. The location intercept effect, $\alpha_{\text{loc}(k)}$, was defined as a normal distribution:

$$\alpha_{\text{loc}(k)} \sim \mathbf{N}(\alpha_{\text{global}}, \tau_{\alpha \text{loc}(k)}) \quad (16)$$

where α_{global} was the global intercept effect and $\tau_{\alpha \text{loc}(k)}$ was the precision attributed to each site (k). The global intercept effect, α_{global} , was defined as a normal distribution:

$$\alpha_{\text{global}} \sim \mathbf{N}(\alpha_0, \tau_{\alpha \text{global}}) \quad (17)$$

where α_0 was the global intercept prior and $\tau_{\alpha \text{global}}$ was the precision of the global intercept. The spatial slope, $\beta_{\text{ind}(jk)}$, was defined as a normal distribution:

$$\beta_{\text{ind}(jk)} \sim \mathbf{N}(\beta_{\text{loc}(k)}, \tau_{\beta \text{ind}}) \quad (18)$$

where $\beta_{\text{loc}(k)}$ was the location slope effect at each site (k) and $\tau_{\beta \text{ind}}$ was the precision of the spatial slope. The location slope effect $\beta_{\text{loc}(k)}$ was defined as a normal distribution:

$$\beta_{\text{loc}(k)} \sim \mathbf{N}(\beta_{\text{global}}, \tau_{\beta \text{loc}(k)}) \quad (19)$$

where β_{global} was the global slope effect and $\tau_{\beta \text{loc}(k)}$ was the precision attributed to each site (k). The global slope effect, β_{global} , was defined as a normal distribution:

$$\beta_{\text{global}} \sim \mathbf{N}(\beta_0, \tau_{\beta \text{global}}) \quad (20)$$

where β_0 was the global slope prior and $\tau_{\beta \text{global}}$ was the precision of the global slope. Species-specific parameters were drawn from normal prior distributions centered at 0 with normal precisions:

$$\alpha_1 \sim \mathbf{N}(0, \tau_{\alpha 1}) \quad (21)$$

The global spatial parameters were drawn from normal prior distributions centered at 0 with normal precisions:

$$\alpha_0 \sim \mathbf{N}(0, \tau_{\alpha 0}) \quad (22)$$

$$\beta_0 \sim \mathbf{N}(0, \tau_{\beta 0}) \quad (23)$$

All precision parameters were given gamma priors because gamma distributions are nonnegative and conjugate to normal distributions. The distributions were:

$$\tau_{\text{obs}}, \tau_{\mu \text{ind}}, \tau_{\mu \text{spp}(l)}, \tau_{\alpha \text{ind}}, \tau_{\alpha \text{loc}(k)}, \tau_{\alpha \text{global}}, \tau_{\beta \text{ind}}, \tau_{\beta \text{loc}(k)}, \tau_{\beta \text{global}}, \tau_{\alpha 1}, \tau_{\alpha 0}, \tau_{\beta 0} \sim \text{gamma}(.01, .1) \quad (24)$$

3.4. Model evaluation

Species-scale GDD thresholds and model performance were evaluated using the posterior distributions of parameters from all four models. To compare differences in model-estimated GDD thresholds for leaf out or budburst, we generated a 500-member posterior distribution of threshold values using the latitude and mean winter temperature for TMA in the spatial models. To compare differences in GDD distributions among models, we used a mixed effects model with either median or 95% credible interval range as the response variable and interactive effects of model (TMA, COMB, LAT, WT) and species group ("TMA + NPN" or "TMA Only") as fixed effects with a random species intercept. Budburst and leaf out were analyzed separately.

Model performance was evaluated using root mean square error (RMSE) and latitudinal trends in model prediction error. For both analyses, we first used the model posterior distributions to calculate a 500-member distribution of GDD threshold values for each budburst or leaf out for USA-NPN observation that was withheld from the model fitting process (Fig. 1). We then used the median GDD value to calculate the day of year of predicted budburst or leaf out for each observation using weather data from Daymet and then calculated the residual difference in days between predicted and observed, from which we calculated RMSE and analyzed latitudinal trends in this residual error. Cross-species model effects on RMSE were analyzed similarly to GDD distribution characteristics with a response variable of RMSE and interactive effects of model and species group and a random species effect.

Latitudinal trends in prediction error were used to infer the ability of each model to capture spatially nonstationary trends in GDD threshold for leaf out and budburst. For this analysis, each species was analyzed independently with prediction error as the response variable, interactive fixed effects of latitude and model, and a random site effect. In the results, we present whether the latitude trend for each result is significantly different from 0 as well as whether there was any trend reduction relative to the single-site TMA model. Degrees of freedom for this analysis vary by species and can be found in Table S5.

All linear mixed effects models used to analyze predictive ability of the phenology models were performed using the lme function in the nlme package in R [46]. All results presented are significant at $\alpha = 0.05$

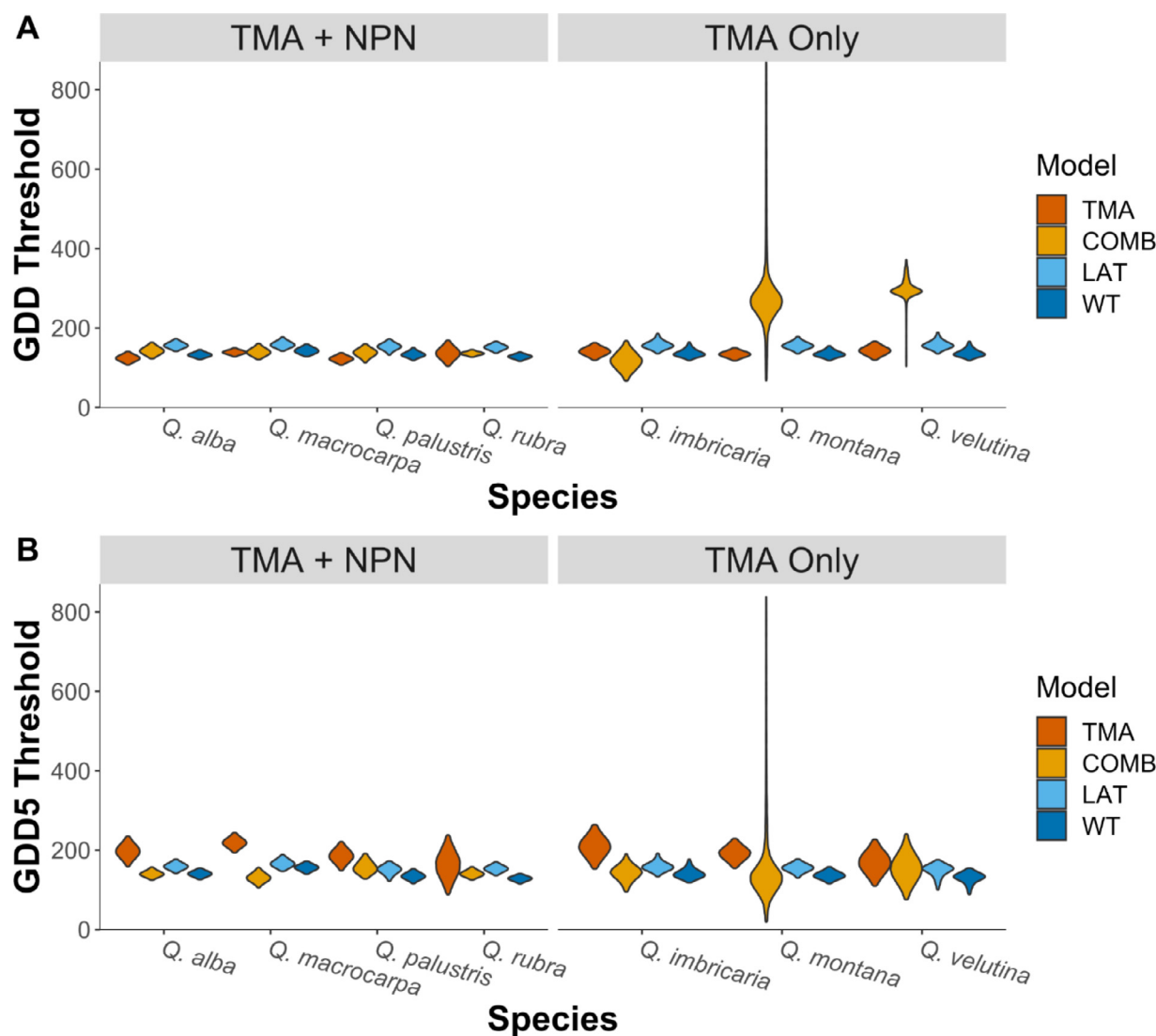


Fig. 2. Model posterior predictions for growing degree-day thresholds (GDD) for budburst (A) and leaf out (B) at The Morton Arboretum (TMA). Shaded areas indicate the 95% credible interval for GDD thresholds with a base temperature of 5 °C for predictions from thermal-time models using TMA data only (TMA), TMA and USA National Phenology Network (USA-NPN) data (COMB), TMA and USA-NPN data with a latitude spatial covariate (LAT), and TMA and USA-NPN data with a mean winter temperature spatial covariate (WT).

with the TMA model as the reference point for comparison for model improvements unless otherwise noted.

4. Results

Inclusion of data from USA-NPN with and without spatial covariates shifted the mean and uncertainty of model-estimated GDD thresholds for budburst and leaf out at The Morton Arboretum (Fig. 2). For budburst, the TMA model had a mean GDD threshold of 134 (SD 8) GDD while the COMB model was statistically higher +42 (SD 71) GDD ($p=0.04$; $df=18$). The GDD thresholds from the LAT and WT models were statistically similar with +23 (SD 8) GDD and WT +0 (SD 8), respectively ($p>0.05$; $df=18$). All leaf GDD thresholds estimated using USA-NPN data were lower than those calculated by the ARB model ($p<0.05$; $df=18$) where the mean ARB threshold was 190 (SD 20 GDD) and model differences were as follows: COMB -48 (SD 26) GDD, LAT -34 (SD 16) GDD, and WT -52 (SD 13) GDD. GDD threshold uncertainty, as measured by the 95% CI range, greatly increased in the COMB model relative to TMA for the “TMA Only” species group for both budburst and leaf out ($p<0.05$; $df=15$) with no difference in range for the LAT or WT models for either

phenophase. Budburst threshold uncertainty increased from 39 (SD 14) GDD in the TMA model to 188 (SD 284) in the COMB model, whereas leaf out increased from 93 (SD 34) GDD to 180 (SD 285), respectively.

In all cases, the inclusion of USA-NPN data with spatial covariates in phenology models reduced budburst and leaf out prediction RMSE (Fig. 3). Model improvements did not vary among “TMA + NPN” and “TMA Only” species groups ($df=15$), but RMSE across all models was higher for the “TMA Only” group for both budburst and leaf out ($df=23$). “TMA + NPN” species had a mean RMSE of 16 (SD 5) days for budburst and 14 (SD 3) days for leaf out, whereas “TMA Only” species had a mean RMSE of 23 (SD 8) days for budburst and 22 (SD 7) days for leaf out. Patterns of RMSE between TMA and COMB models varied among species and phenophases (Fig. 3). When performance was analyzed across all species, there was no statistical difference in RMSE between those two models for budburst ($df=18$), but the COMB model had higher RMSE (3 SD 4 days) for leaf out ($p=0.03$, $df=18$). Both the LAT and WT models displayed significant improvements in model performance as measured by RMSE. Budburst RMSE was reduced by 41% by the LAT model (10 SD 5 days) and 28% by the WT model (7 SD 3 days) compared to the TMA model. RMSE improvements in the two

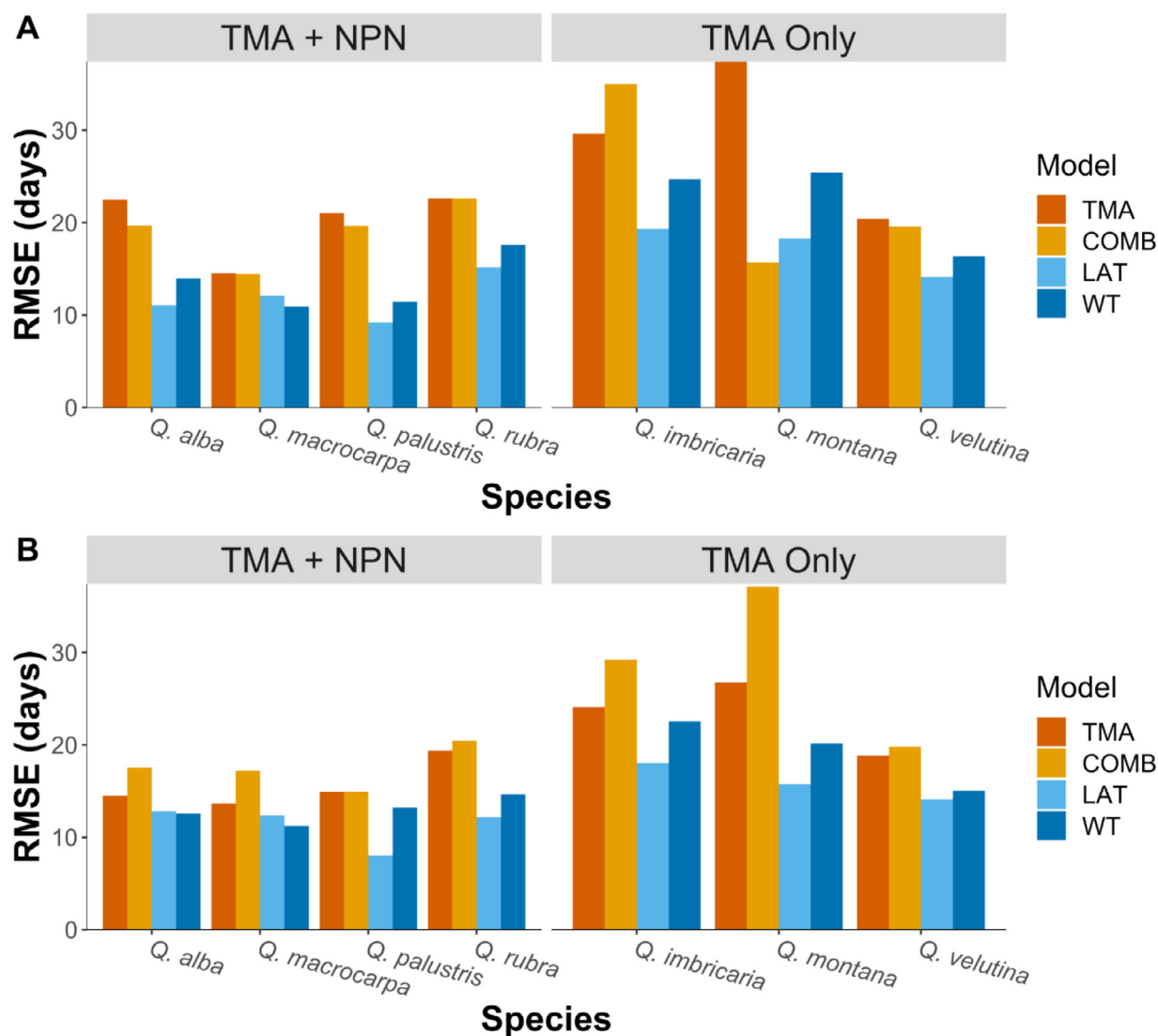


Fig. 3. (in color): Root mean square error (RMSE, in days) for predicted budburst (A) and leaf out (B) from thermal-time models using The Morton Arboretum data only (TMA), TMA and USA National Phenology Network (USA-NPN) data (COMB), TMA and USA-NPN data with a latitude spatial covariate (LAT), and TMA and USA-NPN data with a mean winter temperature spatial covariate (WT). For species whose GDD thresholds were modeled using TMA and USA-NPN data (“TMA + NPN”), RMSE was calculated on a random subset of 25% of USA-NPN sites excluded from model calibration whereas RMSE for “TMA Only” species were calculated using all available USA-NPN observations.

spatial models were more modest for leaf out, which had overall lower RMSE than budburst in the TMA model (19 SD 5 days for leaf out versus 24 SD 7 days for budburst). For leaf out, RMSE was reduced by 29% (6 SD 3 days) in the LAT model and 17% (3 SD 2 days) in the WT model.

There was a significant latitudinal trend in the single-site TMA model error for six of seven species for budburst and five species for leaf out (Fig. 4, Table S5). Across all species, the mean latitudinal trend in model error for the TMA model was -3.6 (SD 2.2) days per degree latitude for budburst and -3.4 (SD 2.1) days per degree latitude for leaf out, indicating that models predicted later budburst than observed at southern latitudes and earlier than observed at northern sites. No statistically significant latitudinal trend was found in any models or phenophase predictions for *Q. montana*, which had a relatively small latitudinal distribution compared to the other species (Fig. 1, Fig. 4). Furthermore, the addition of data from USA-NPN alone, without a spatial covariate, only reduced the trend in prediction error for *Q. velutina* budburst, which had a statistically significant trend of -2.3 (SE 1.1) days per degree latitude in the TMA model to a nonsignificant trend of -1.1 (SE 1.1) days per degree latitude in the COMB model (Table S5). Of the “TMA Only” species, only *Q. velutina*, which had the largest

latitudinal range (Fig. 1, Fig. 4), showed significant improvement in latitudinal error trends in both LAT and WT spatial covariate models (Table S5). However, the LAT and WT models reduced latitudinal error trends in all “TMA + NPN” species (Fig. 4, Table S5). Across all species, the LAT model reduced latitudinal trends by a mean of 3.2 (SD 0.1) days per degree latitude for budburst and 3.4 (SD 0.3) days per degree latitude for leaf out. The WT model showed less improvement than LAT with mean trend reductions of 1.3 (SD 0.7) days per degree latitude for budburst and 1.8 (SD 0.8) days per degree latitude for leaf out.

5. Discussion

Living botanical collections and data originating from citizen science networks offer complementary strengths that can be combined to improve phenology predictions for data-poor species, including rare species whose vulnerability to climate change is often difficult to study. Phenology monitoring programs at botanical gardens and arboreta are often able to observe hundreds of taxa from across the world, including rare and data-poor species that have been collected specifically for *ex situ* study and conservation [25,47]. In our study, we

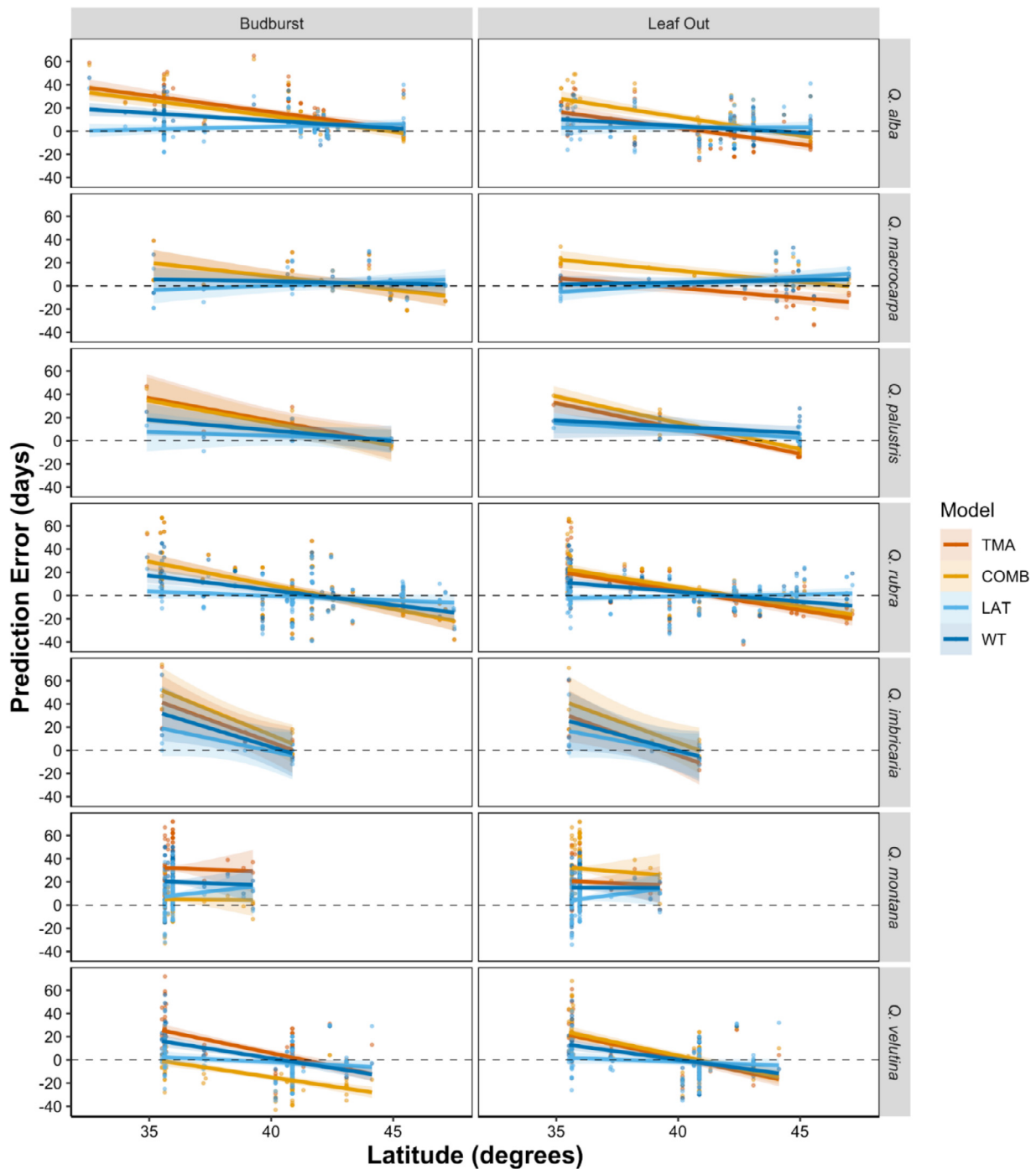


Fig. 4. (in color): Latitudinal trends in model prediction error from four thermal-time models of leaf out and budburst: The Morton Arboretum data only (TMA), TMA and USA National Phenology Network (USA-NPN) data (COMB), TMA and USA-NPN data with a latitude spatial covariate (LAT), and TMA and USA-NPN data with a mean winter temperature spatial covariate (WT). Points indicate the difference between predicted and observed for USA-NPN observations not used in the model fitting process. For the four “TMA + NPN” species (*Q. alba*, *Q. macrocarpa*, *Q. palustris*, *Q. rubra*) points and latitudinal trends shown are for a random subset of 25% of sites available in the USA-NPN datasets. The three remaining species (*Q. imbricaria*, *Q. montana*, *Q. velutina*) only used data from The Morton Arboretum in the model fitting process and all available USA-NPN data was used for model evaluation.

have leveraged this diversity by jointly modeling the GDD thresholds for leaf out and budburst for multiple congeneric species. However, relatively high RMSE and strong latitudinal trends from our TMA model compared to others demonstrates the common limitation of single-site studies in being able to predict responses outside of the limited range of conditions captured in the training data [48,49]. Conversely, the geographically dispersed nature of citizen science datasets leads them to better capture the observations for individual species throughout

their native ranges than is possible with datasets focused on a single or even a few intensive sites (Fig. 1). The broad geographic distribution of these datasets is often mirrored by broad representation of organismal types and phylogenetic clades, although individual clades or functional groups such as *Quercus* may be sparsely represented relative to their global taxonomic richness [34,50].

The ability of large, spatially distributed datasets, including those from citizen science, to improve phenological predictions is not inherent

and additional steps to describe other sources of variability in observations is essential for improved model performance. The noise introduced into models from spatially broad citizen science data without the inclusion of spatial covariates can increase model parameter uncertainty (Fig. 2) as well as overall error in many instances (Fig. 3). The lack of consistent improvements in COMB model performance, as measured by RMSE (Fig. 3) and latitudinal error trends (Fig. 4), indicate that the power of citizen science phenology networks, such as USA-NPN, stems from its ability to allow for more complex model fitting that captures spatio-temporal variation in phenological patterns rather than through an intrinsic power of adding more data. Our results demonstrate that consideration of spatial nonstationarity in phenological cues [48,51] and the incorporation of data able to describe those patterns are essential for accurate inference and prediction of biological responses to climate variability and change over large areas (Fig. 3, Fig. 4).

As climate change threatens global biodiversity and the particular challenges of long-lived organisms such as trees increase in their immediacy [52,53], arboreta and botanical gardens provide critical opportunities to study the biology and climate responses of rare, endemic, and data-poor species that are often infeasible to study in their natural habitats. This study restricted its analyses to taxa with robust data available through USA-NPN for prediction evaluation, but botanical gardens often include the opportunity to compare dozens or hundreds of taxa commonly under-represented in traditional “big data” networks [22,23], and particularly networks that require repeat observations such as USA-NPN [14]. For example, contribution of TMA observations to the USA-NPN creates a 50% increase in the number of *Q. palustris* trees observed in a national-scale dataset (Tables S1, S2). However, data from a single location alone is often limited in its ability to describe species’ responses to weather over a wide spatio-temporal range of climate variability [30,31,48], and thus coordination and participation with larger networks remains essential to phenology and climate change research.

Large scale citizen science networks such as USA-NPN are essential for climate change research and conservation as one of the few approaches for gathering individual and species-scale data over continental geographic extents. Data maintained by the USA-NPN have been used to make species-level forecasts of activity for dozens of both plant and animal species to be used for both research and management-focused audiences [15–17]. However, only a fraction of species present in citizen science programs such as USA-NPN have sufficient data density to parameterize robust phenological models [54]. We used a relatively simple thermal time phenological model in our study because the “TMA Only” species group that was representative of sparsely observed or rare species did not have sufficient data to parameterize multiple species-level phenological cues such as winter chilling or daylength requirements [27,28]. A more complex and explicit representation of the biological or ecological relationships among species from TMA and those in USA-NPN based on traits or phylogenetic relatedness may provide continued pathways for improvement [10,32,55,56]. However, when spatially broad citizen science data are used without attempts to capture spatial nonstationarity, the additional observations do not improve predictive ability. This lack of improvement was demonstrated in our study where model RMSE and latitudinal trends in prediction error were similar between the TMA model parameterized with only two years of data from a single site and the COMB model using both TMA and USA-NPN data (Fig. 3, Fig. 4). In fact, without additional measures, undescribed sources of variation within sparse data can create additional challenges for model fitting, as seen with *Q. montana* in the COMB model (Fig. 3). In this instance, TMA is not within the native range of *Q. montana* and has the greatest difference in latitude between TMA and USA-NPN observations (Fig. 1, Tables S1, S2). The lack of spatial overlap between TMA observations made outside *Q. montana*’s native range and the USA-NPN validation observations increased uncertainty in latitudinal trends, leading to a lack of improvement in spatial trends in phenology prediction (Fig. 4, Table S5), even though *Q. montana* had a similarly reduced RMSE as the other species in this study (Fig. 3).

Thus, while hierarchical modeling is a powerful approach to improve phenology predictions for rare or data-poor species, those with high unexplained variation will remain challenging without more robust data.

Phenology predictions for rare and sparsely observed species can be improved by combining information from single sites, including botanical living collections outside of a species range, with observations of similar taxa in geographically broad citizen science networks. Ideally, observations that capture the full environmental diversity of each species of interest would be available for predicting responses to current and future global change, but the logistical challenges of obtaining individual- and species-scale data for rare species make this an unrealistic goal. “Big data” approaches for characterizing phenology such as citizen science and remote sensing have the potential to partially address spatial variation in phenology at continental and global extents [57–59]. However, the ecology of many species with restricted distributions, including those of conservation concern, remain poorly described and unaddressed through “big data” [14]. Living collections monitoring programs reduce logistical hurdles for observing rare species and provide access necessary for the frequent observations needed to describe phenology for individual species. Using hierarchical modeling approaches that combine detailed species observations from even a single site with comparable data for related or ecologically similar taxa that capture spatial variability can be a powerful tool for borrowing strength across research approaches. Through a combination of two approaches to phenological monitoring with known limitations, botanical garden observations and citizen science observations can be combined to provide key information necessary to better understand and predict the responses of rare and sparsely-observed species to climate variability and change across space.

Acknowledgements

We thank the volunteers of The Morton Arboretum’s phenology program for providing the data that makes this research possible. Data were provided by the USA National Phenology Network and the many participants who contribute to its *Nature’s Notebook* program. We thank the USA National phenology network for its ongoing partnership and are appreciative of the efforts of all USA-NPN contributors. A.S. was funded through a University of Chicago Empower Scholarship as a part of the Jeff Metcalf Internship Program. We also thank E. Wolkovich and three anonymous reviewers whose highly constructive comments greatly strengthened this manuscript.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:[10.1016/j.ecoach.2021.100032](https://doi.org/10.1016/j.ecoach.2021.100032).

References

- [1] C. Parmesan, M.E. Hanley, Plants and climate change: complexities and surprises, *Ann. Bot.* 116 (2015) 849–864, doi:[10.1093/aob/mcv169](https://doi.org/10.1093/aob/mcv169).
- [2] A. Menzel, T.H. Sparks, N. Estrella, E. Koch, A. Aasa, R. Ahas, K. Alm-Kübler, P. Bissolli, O. Braslavská, A. Briede, F.M. Chmielewski, Z. Crepinsek, Y. Curnel, Å. Dahl, C. Defila, A. Donnelly, Y. Filella, K. Jatzcak, F. Mège, A. Mestre, Ø. Nordli, J. Peñuelas, P. Pirinen, V. Remišová, H. Scheffinger, M. Striz, A. Susnik, A.J.H. Van Vliet, F.-E. Wielgolaski, S. Zach, A. Züst, European phenological response to climate change matches the warming pattern, *Glob. Change Biol.* 12 (2006) 1969–1976, doi:[10.1111/j.1365-2486.2006.01193.x](https://doi.org/10.1111/j.1365-2486.2006.01193.x).
- [3] I. Chuine, Why does phenology drive species distribution? *Philos. Trans. R. Soc. B Biol. Sci.* 365 (2010) 3149–3160, doi:[10.1098/rstb.2010.0142](https://doi.org/10.1098/rstb.2010.0142).
- [4] A. Duputié, A. Rutschmann, O. Ronce, I. Chuine, Phenological plasticity will not help all species adapt to climate change, *Glob. Change Biol.* 21 (2015) 3062–3073, doi:[10.1111/gcb.12914](https://doi.org/10.1111/gcb.12914).
- [5] L.J. Martin, B. Blossey, E. Ellis, Mapping where ecologists work: biases in the global distribution of terrestrial ecological observations, *Front. Ecol. Environ.* 10 (2012) 195–201, doi:[10.1890/110154](https://doi.org/10.1890/110154).
- [6] M. Pacifici, W.B. Foden, P. Visconti, J.E.M. Watson, S.H.M. Butchart, K.M. Kovacs, B.R. Scheffers, D.G. Hole, T.G. Martin, H.R. Akçakaya, R.T. Corlett, B. Huntley, D. Bickford, J.A. Carr, A.A. Hoffmann, G.F. Midgley, P. Pearce-Kelly, R.G. Pearson, S.E. Williams, S.G. Willis, B. Young, C. Rondinini, Assessing species vulnerability to climate change, *Nat. Clim. Change.* 5 (2015) 215–224, doi:[10.1038/nclimate2448](https://doi.org/10.1038/nclimate2448).

- [7] K. Havens, P. Vitt, J. Schwarz, B. Orr, T. Crimmins, Chicago Botanic Garden's conservation and outreach efforts on climate change, *BGjournal* 4 (2007) 13–16.
- [8] I. Lacher, M.W. Schwartz, Empirical test on the relative climatic sensitivity between individuals of narrowly and broadly distributed species, *Ecosphere* 7 (2016), doi:10.1002/ecs2.1227.
- [9] S.M. Still, A.L. Frances, A.C. Treher, L. Oliver, Using two climate change vulnerability assessment methods to prioritize and manage rare plants: a case study, *Nat. Areas J.* 35 (2015) 106–121, doi:10.3375/043.035.0115.
- [10] M.W. Cadotte, J.T. Davies, Rarest of the rare: advances in combining evolutionary distinctiveness and scarcity to inform conservation at biogeographical scales: conservation phylo-biogeography, *Divers. Distrib.* 16 (2010) 376–385, doi:10.1111/j.1472-4642.2010.00650.x.
- [11] L.R. Prugh, A.R.E. Sinclair, K.E. Hodges, A.L. Jacob, D.S. Wilcoxe, Reducing threats to species: threat reversibility and links to industry: reducing threats to species, *Conserv. Lett.* 3 (2010) 267–276, doi:10.1111/j.1755-263X.2010.00111.x.
- [12] J.R. Bennett, G. Elliott, B. Mellish, L.N. Joseph, A.I.T. Tulloch, W.J.M. Probert, M.M.I. Di Fonzo, J.M. Monks, H.P. Possingham, R. Maloney, Balancing phylogenetic diversity and species numbers in conservation prioritization, using a case study of threatened species in New Zealand, *Biol. Conserv.* 174 (2014) 47–54, doi:10.1016/j.biocon.2014.03.013.
- [13] H.K. Kindsvater, N.K. Dulvy, C. Horswill, M.-J. Juan-Jordá, M. Mangel, J. Matthiopoulos, Overcoming the data crisis in biodiversity conservation, *Trends Ecol. Evol.* 33 (2018) 676–688, doi:10.1016/j.tree.2018.06.004.
- [14] J.L. Dickinson, B. Zuckerman, D.N. Bonter, Citizen Science as an ecological research tool: challenges and benefits, *Ann. Rev. Ecol. Syst.* 41 (2010) 149–172, doi:10.1146/annurev-ecolsys-102209-144636.
- [15] I. Ibáñez, R.B. Primack, A.J. Miller-Rushing, E. Ellwood, H. Higuchi, S.D. Lee, H. Kobori, J.A. Silander, Forecasting phenology under global warming, *Philos. Trans. R. Soc. B Biol. Sci.* 365 (2010) 3247–3260, doi:10.1098/rstb.2010.0120.
- [16] S.D. Taylor, E.P. White, Automated data-intensive forecasting of plant phenology throughout the United States, *Ecol. Appl.* (2020) 30, doi:10.1002/eap.2025.
- [17] T.M. Crimmins, K.L. Gerst, D.G. Huerta, R.L. Marsh, E.E. Posthumus, A.H. Rosemartin, J. Switzer, J.F. Weltzin, L. Coop, N. Dietschler, D.A. Herms, S. Limbu, R.T. Trotter, M. Whitmore, Short-term forecasts of insect phenology inform pest management, *Ann. Entomol. Soc. Am.* 113 (2020) 139–148, doi:10.1093/aesa/saz026.
- [18] D. Jerome, E. Beckman, L. Kenny, K. Wenzel, C.-S. Kua, M. Westwood, *The Red List of US Oaks*, The Morton Arboretum, Lisle, IL, 2017.
- [19] R. Engler, A. Guisan, L. Rechsteiner, An improved approach for predicting the distribution of rare and endangered species from occurrence and pseudo-absence data, *J. Appl. Ecol.* 41 (2004) 263–274, doi:10.1111/j.0021-8901.2004.00881.x.
- [20] A. Guisan, A. Lehmann, S. Ferrier, M. Austin, J.M.C. Overton, R. Aspinall, T. Hastie, Making better biogeographical predictions of species' distributions, *J. Appl. Ecol.* 43 (2006) 386–392, doi:10.1111/j.1365-2664.2006.01164.x.
- [21] C.C. Davis, C.G. Willis, B. Connolly, C. Kelly, A.M. Ellison, Herbarium records are reliable sources of phenological change driven by climate and provide novel insights into species' phenological cueing mechanisms, *Am. J. Bot.* 102 (2015) 1599–1609, doi:10.3732/ajb.1500237.
- [22] L. Faraji, M. Karimi, Botanical gardens as valuable resources in plant sciences, *Biodivers. Conserv.* (2020), doi:10.1007/s10531-019-01926-1.
- [23] R.B. Primack, E.R. Ellwood, A.S. Gallinat, A.J. Miller-Rushing, The growing and vital role of botanical gardens in climate change research, *New Phytol.* 231 (2021) 917–932, doi:10.1111/nph.17410.
- [24] I.S. Pearse, A.L. Hipp, Global patterns of leaf defenses in oak species, *Evolution* 66 (2012) 2272–2286, doi:10.1111/j.1558-5646.2012.01591.x.
- [25] Z.A. Panchen, R.B. Primack, B. Nordt, E.R. Ellwood, A.-D. Stevens, S.S. Renner, C.G. Willis, R. Fahey, A. Whittemore, Y. Du, C.C. Davis, Leaf out times of temperate woody plants are related to phylogeny, deciduousness, growth habit and wood anatomy, *New Phytol.* 203 (2014) 1208–1219, doi:10.1111/nph.12892.
- [26] Z.A. Panchen, R.B. Primack, A.S. Gallinat, B. Nordt, A.-D. Stevens, Y. Du, R. Fahey, Substantial variation in leaf senescence times among 1360 temperate woody plant species: implications for phenology and ecosystem processes, *Ann. Bot.* 116 (2015) 865–873, doi:10.1093/aob/mcv015.
- [27] D. Basler, Evaluating phenological models for the prediction of leaf-out dates in six temperate tree species across central Europe, *Agric. For. Meteorol.* 217 (2016) 10–21, doi:10.1016/j.agrformet.2015.11.007.
- [28] K. Hufkens, D. Basler, T. Milliman, E.K. Melaas, A.D. Richardson, An integrated phenology modelling framework in R, *Methods Ecol. Evol.* 9 (2018) 1276–1285, doi:10.1111/2041-210X.12970.
- [29] S.C. Elmendorf, T.M. Crimmins, K.L. Gerst, J.F. Weltzin, Time to branch out? Application of hierarchical survival models in plant phenology, *Agric. For. Meteorol.* 279 (2019) 107694, doi:10.1016/j.agrformet.2019.107694.
- [30] C. Olsson, A.M. Jönsson, Process-based models not always better than empirical models for simulating budburst of Norway spruce and birch in Europe, *Glob. Change Biol.* 20 (2014) 3492–3507, doi:10.1111/gcb.12593.
- [31] S.D. Taylor, J.M. Meiners, K. Riemer, M.C. Orr, E.P. White, Comparison of large-scale citizen science data and long-term study data for phenology modeling, *Ecology* 100 (2019) e02568, doi:10.1002/ecy.2568.
- [32] S.J. Mazer, S.E. Travers, B.I. Cook, T.J. Davies, K. Bolmgren, N.J.B. Kraft, N. Salamin, D.W. Inouye, Flowering date of taxonomic families predicts phenological sensitivity to temperature: Implications for forecasting the effects of climate change on unstudied taxa, *Am. J. Bot.* 100 (2013) 1381–1397, doi:10.3732/ajb.1200455.
- [33] E.G. Denny, K.L. Gerst, A.J. Miller-Rushing, G.L. Tierney, T.M. Crimmins, C.A.F. Enquist, P. Guertin, A.H. Rosemartin, M.D. Schwartz, K.A. Thomas, J.F. Weltzin, Standardized phenology monitoring methods to track plant and animal activity for science and resource management applications, *Int. J. Biometeorol.* 58 (2014) 591–601, doi:10.1007/s00484-014-0789-5.
- [34] Rosemartin, A., Denny, E.G., Gerst, K.L., Marsh, R.L., Posthumus, E.E., Crimmins, T.M., and Weltzin, J.F., 2018, USA National Phenology Network observational data documentation: U.S. Geological Survey Open-File Report 2018–1060, 24 p., doi:10.3133/ofr20181060.
- [35] S. Chamberlain, L. Marsh, A. Rosemartin, K. Wong, rnpn:Interface to the National 'Phenology' Network 'API'. 2020
- [36] K.L. Gerst, N.L. Rossington, S.J. Mazer, Phenological responsiveness to climate differs among four species of *Quercus* in North America, *J. Ecol.* 105 (2017) 1610–1622, doi:10.1111/1365-2745.12774.
- [37] Little EL (1971) Atlas of United States Trees (US Department of Agriculture, Forest Service, Washington, DC).
- [38] K. Hufkens, D. Basler, T. Milliman, E. K. Melaas, A. D. Richardson, An integrated phenology modelling framework in R: modelling vegetation phenology with phenor, *Methods Ecol. Evol.* 9 (2) (2018) 1–10.
- [39] P.E. Thornton, S.W. Running, M.A. White, Generating surfaces of daily meteorological variables over large regions of complex terrain, *J. Hydrol.* 190 (1997) 214–251, doi:10.1016/S0022-1694(96)03128-9.
- [40] [dataset] M.M. Thornton, R. Shrestha, Y. Wei, P.E. Thornton, S. Kao, and B.E. Wilson. Daymet: Daily Surface Weather Data on a 1-km Grid for North America, Version 4. ORNL DAAC, Oak Ridge, Tennessee, USA. 2020. doi:10.3334/ORNLDAAC/1840.
- [41] M. Plummer, JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling, (2003).
- [42] M. Plummer, rjags: Bayesian Graphical Models using MCMC. <https://CRAN.R-project.org/package=rjags>
- [43] M. Plummer, N. Best, K. Cowles, K. Vines, CODA: Convergence Diagnosis and Output Analysis for MCMC, *R News* 6 (2006) 7–11.
- [44] R Core Team: R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, Austria, 2020 <https://www.R-project.org>.
- [45] RstudioStudio: Integrated development for R, RStudio, PBC, Boston, MA, 2020 <http://www.rstudio.com>.
- [46] J. Pinheiro, D. Bates, S. DebRoy, D. Sarkar, R Core Team, 2020 <https://CRAN.R-project.org/package=nlme>.
- [47] R. Primack, I. Ibáñez, H. Higuchi, S. Lee, A.J. Miller-Rushing, A.J. Miller-Rushing, A.M. Wilson, J. Silander, Spatial and interspecific variability in phenological responses to warming temperatures, *Biol. Conserv.* (2009) 2569–2577, doi:10.1016/j.biocon.2009.06.003.
- [48] C.R. Rollinson, A.O. Finley, M.R. Alexander, S. Banerjee, K.-A. Dixon Hamil, L. Koenig, D. Locke, M. Peterson, M. Tingley, K. Wheeler, C. Youngflesh, E. Zipkin, Working across space and time: nonstationarity in ecological research and application, *Front. Ecol. Environ.* (2021).
- [49] D. Basler, Evaluating phenological models for the prediction of leaf-out dates in six temperate tree species across central Europe, *Agric. For. Meteorol.* 217 (2016) 10–21, doi:10.1016/j.agrformet.2015.11.007.
- [50] [dataset] USA National Phenology Network, Plant and Animal Phenology Data., Data Type Site Phenometrics. (2020) 2008–12/31/2019 for Region: 48.0459°, -67.7731° (UR); 26.9280°, -123.3737° (LL). USA-NPN, Tucson, Arizona, USA. Data set accessed 03/21/2019 at doi:10.5066/F7854N1.
- [51] E.M. Wolkovich, B.I. Cook, K.K. McLauchlan, T.J. Davies, Temporal ecology in the Anthropocene, *Ecol. Lett.* 17 (2014) 1365–1379, doi:10.1111/ele.12353.
- [52] S.N. Aitken, S. Yeaman, J.A. Holliday, T. Wang, S. Curtis-McLane, Adaptation, migration or extirpation: climate change outcomes for tree populations, *Evol. Appl.* 1 (2008) 95–111, doi:10.1111/j.1752-4571.2007.00013.x.
- [53] C.D. Allen, A.K. Macalady, H. Chenchouni, D. Bachelet, N. McDowell, M. Vennetier, T. Kitzberger, A. Rigling, D.D. Breshears, E.H. (Ted) Hogg, P. Gonzalez, R. Fensham, Z. Zhang, J. Castro, N. Demidova, J.-H. Lim, G. Allard, S.W. Running, A. Semerci, N. Cobb, A global overview of drought and heat-induced tree mortality reveals emerging climate change risks for forests, *For. Ecol. Manag.* 259 (2010) 660–684, doi:10.1016/j.foreco.2009.09.001.
- [54] T.M. Crimmins, M.A. Crimmins, K.L. Gerst, A.H. Rosemartin, J.F. Weltzin, USA National Phenology Network's volunteer-contributed observations yield predictive models of phenological transitions, *PLOS ONE* 12 (2017) e0182919, doi:10.1371/journal.pone.0182919.
- [55] J. Cavender-Bares, K.H. Kozak, P.V.A. Fine, S.W. Kembel, The merging of community ecology and phylogenetic biology, *Ecol. Lett.* 12 (2009) 693–715, doi:10.1111/j.1461-0248.2009.01314.x.
- [56] D. Li, A.R. Ives, D.M. Waller, Can functional traits account for phylogenetic signal in community composition? *New Phytol.* 214 (2017) 607–618, doi:10.1111/nph.14397.
- [57] E.K. Melaas, M.A. Friedl, A.D. Richardson, Multiscale modeling of spring phenology across Deciduous Forests in the Eastern United States, *Glob. Change Biol.* 22 (2016) 792–805, doi:10.1111/gcb.13122.
- [58] J. Tang, C. Körner, H. Muraoka, S. Piao, M. Shen, S.J. Thackeray, X. Yang, Emerging opportunities and challenges in phenology: a review, *Ecosphere* 7 (2016), doi:10.1002/ecs2.1436.
- [59] A.D. Richardson, K. Hufkens, T. Milliman, D.M. Aubrecht, M. Chen, J.M. Gray, M.R. Johnston, T.F. Keenan, S.T. Klosterman, M. Kosmala, E.K. Melaas, M.A. Friedl, S. Froliking, Tracking vegetation phenology across diverse North American biomes using PhenoCam imagery, *Sci. Data.* 5 (2018) 180028, doi:10.1038/sdata.2018.28.