Estimating flowering transition dates from status-based phenological observations: a test of methods

Shawn D. Taylor[1]

[1] School of Natural Resources and Environment, University of Florida Gainesville, FL, United States

Abstract

The scale of phenological research has expanded due to the digitization of herbarium specimens and volunteer based contributions. These data are status-based, representing the presence or absence of a specific phenophase. Modelling the progress of plant dormancy to growth and reproduction and back to dormancy requires estimating the transition dates from these status-based observations. There are several methods available for this ranging from statistical moments using the day of year to newly introduced methods using concepts from other fields. Comparing the proficiency of different estimators is difficult since true transition dates are rarely known. Here I use a recently released dataset of in-situ flowering observations of the perennial forb *Echinacea angustifolia.* In this dataset, due to high sampling frequency and unique physiology, the transition dates of onset, peak, and end of flowering are known to within 3 days. I used a Monte Carlo analysis to test eight different estimators across two scales using a range of sample sizes and proportion of flowering presence observations. I evaluated the estimators accuracy in predicting the onset, peak, and end of flowering at the population level, and predicting onset and end of flowering for individual plants. Overall a method using a Weibull distribution performed the best for population level onset and end estimates, but other estimators may be more appropriate when there is a large amount of absence observations relative to presence observations. For individual estimates a method using the midway point between the first flower presence and most prior flower absence, within 7 days, is the best option as long as the restriction does not limit the final sample size. Otherwise the Weibull method is adequate for individual estimates as well. These methods allow practitioners to effectively utilize the large amount of status-based phenological observations currently available.

Keywords: onset, peak, flowering, budburst, sampling frequency, herbarium records, GAM

Introduction

Plant phenology has a long history in ecological research and is a primary indicator of climate change (Scheffers et al. 2016, Chuine and Régnière 2017). Studies commonly document the long-term trends of the first flower or leaf out dates, apply various modelling approaches to infer the drivers of these transitions, or make forecasts using future climate conditions. Phenological models, such as predictive models or those used for long-term trends, use the transition dates as the variable of interest. Common transition dates are First Observed open flowers or new leaves on a plant, but can also include peak flower, fruit maturation, and leaf senescence. Historic datasets often use repeated observations to identify the true transition date (Wolkovich et al. 2012, Davis et al. 2015), yet this is susceptible to observer bias (Miller-Rushing et al. 2008). Most modern studies and collection protocols use status-based monitoring, where over time observers record the current state of a single plant (ie. leaves present or absent) regardless of recent or impending transitions. This includes research using herbarium records, where the presence or absence of flowers and other phenophases is inferred from their presence on a specimen (Willis et al. 2017). To make use of status-based data in most phenological models the transition date must first be estimated, and there are several methods available.

Two of the most common estimators are the First Observed and Mean Flowering methods, where either the first observation in a year or the mean dates for all observations within a year is used as an estimate for a phenophase transition and peak dates, respectively (Miller-Rushing et al. 2008, CaraDonna et al. 2014, Willis et al. 2017, Jones and Daehler 2018). The First Observed method has been shown to be biased in several instances, while the Mean Flowering date is considered a reliable estimator for the midpoint or peak of a phenophase (Miller-Rushing et al. 2008, Moussus et al. 2010, Bertin 2015). Recently more robust methods have been introduced. Templ et al. (2017) used survival modelling to estimate the median date of flowering

60 and Pearse et al. (2017) used an extinction model to estimate the first flowering date.

61 Using repeated observations of individual plants, as opposed to observations from

62 across a population, site, or region, allows for more reliable estimates. For example if

63 flowers are not present during one visit but present during the next, the transition of

64 flowers opening is constrained to the window between the two visits (Gerst et al. 2016).

65 Studies of bird migration phenology face similar challenges and several estimators have

66 been used to model the first arrival dates. Examples include logistic regression (Mayor

67 et al. 2017) and General Additive Models (GAMs) (Moussus et al. 2009, Newson et al.

68 2016, Lindén et al. 2017). To date no comparison has been made of these different

69 transition date estimators for plant phenology.

70 Furthermore, there are no clear guidelines for using estimators across different spatial

71 scales. Over a latitudinal gradient the transition of a phenophase for a single species can

72 last several weeks to months, and even at the local scale can vary due to many factors

73 (Diez et al. 2012, Zhang et al. 2017). Studies which estimate transition dates have

74 combined observations from individual plants (Gerst et al. 2016, Taylor et al. 2019),

75 sites or populations of plants (Schaber and Badeck 2002, Linkosalo et al. 2008, Basler

76 2016), or entire regions (Calinger et al. 2013, Park 2014). How different phenological

77 estimators perform across spatial scales is currently unknown.

78 A comparison of estimators is difficult since, due to infrequent sampling, the true date

79 of transitions is rarely known. Previously Moussus et al. (2010) used simulated data to

80 test the ability of different estimators to detect shifts in phenological distributions. Here

81 I expand on this prior study by using a dataset of flowering observations from a single

82 population where, due to the unique physiology of the focal species, transition dates can

83 be calculated with high precision, and the efficacy of the different estimators directly

84 compared. To determine how these estimators perform using different sources of

85 phenological data, such as those from herbarium records or crowd-source applications, I

86 performed this analysis across two different scales (population and individual level

87 transition dates), with varying sample sizes, and with varying proportions of observed

88 flowering presence.

Methods

Phenological Data

91 I used phenological observations of the perennial forb *Echinacea angustifolia* collected

92 in Minnesota, U.S.A. in the years 1995-2015 (Waananen et al. 2018a, 2018b) to test the

93 accuracy of different transition date estimators. The data consist of the start and end

94 date of flowering (defined as the start and end of pollen production) for 286 individual

95 plants in a 0.5 ha plot from the 11 years of sampling, where the sampling frequency was

96 at least every 3 days during pollen production. The flowering of *E. angustifolia* is such

97 that the true start date of flowering can be inferred very precisely for an individual plant.

98 The flowering heads of *E. angustifolia* consist of 80-250 disk flowers in several rows.

99 The bottom most row flowers first, with each adjacent row flowering every day

100 afterwards. This pattern was used to determine the date of first flowering for an

101 individual to within 2 days for flowering onset and 3 days for flowering end (Wagenius

102 2004, Waananen et al. 2018b). With this information a true start and end date of

103 flowering for the entire population can be approximated.

104 Different interpretations of phenological metrics can yield different results (Renzi et al.

105 2019), thus with the *E. angustifolia* dataset I used strict definitions in calculating the

106 true values used in the analysis. For each year I calculated the following

107 population-level metrics: 1) the start of flowering as defined by the day of year (DOY)

108 of the first observed flower, 2) peak flower defined as the DOY when the most flowers

109 were observed in a given year, 3) the end of flowering as defined by the last DOY a

110 flower was observed. I also calculated two individual level metrics: 1) the start and 2)

111 end DOY of flowering for each individual plant in each year.

112 To simulate status-based data of a plant population I first determined the flowering

113 status (either present or absent) for every individual plant on every DOY 1-365, then

114 randomly sampled from these dates. Thus an observation could be of flowers present or

115 absent. Flowering absence observations are possible throughout the year as no

116 individual flowers for the full duration of the season. I performed a Monte Carlo

117 analysis, where for every year I repeated this 1000 times with varying sample sizes (10,

118 50, and 100 observations) and varying levels of flowering presence being observed

119 (25%, 50%, and 75%). For example, with flowering presence set to 25% using 100

120 observations only 25 observations were allowed to be of flower presence while the rest

121 were of flower absence, all being randomly chosen from the full calendar year. The

122 variation in sample size and ratio of flowering presence observations simulate patterns

123 seen in non-systematic phenological datasets, such as the those from herbarium records

124 or volunteer contributions. These patterns stem from biases in the time of year of

125 sampling, infrequent or sporadic sampling, or variations in observer effort (Dickinson et

126 al. 2010, Willis et al. 2017, Daru et al. 2018).

127 For individual level flowering estimates I performed the same random sampling routine

128 for every individual in every year using sample sizes of 5, 10, and 20 observations, and

129 flowering presence ratios of 25%, 50%, and 75%. I only used individuals which were in

130 flower for more than 20 days, since below that there would not be enough data in the

131 lowest sample size and flowering presence classes. I repeated this 20 times for each

132 individual in every year in the Monte Carlo analysis.

133 Each estimator, described below, was fit to each random sample (Table 1). For the

134 population estimates this resulted in 11,000 estimates for each estimator, and sample

135 size/flowering presence combination. For the individual estimates this resulted in 4840

136 estimates for each estimator and sample size/flowering presence combination.

137 Estimators were compared using the R^2 value between estimated and observed dates of

138   metrics, and by examining the density of errors from all Monte Carlo estimates.

Estimators

140   The First Observed method uses the earliest DOY of flowering as the estimate for the

141   start of flowering. Analogous to this is the Last Observed DOY for estimating the end of

142   flowering. These were used in both the population and individual level analysis.

143   The Midway method uses the midway date between the First Observed flowering date

144   and the most prior observation of flowering absence for an individual plant. This can be

145   improved by applying a restriction whereas only individuals with an observed absence

146   within 7 days of the First Observed presence are used (Gerst et al. 2016). Applying this

147   restriction reduces the final sample size available for modelling though. The Midway

148   method was used to estimate onset and end in the individual analysis (Midway and

149   Midway 7-Day), and in the population analysis by using the mean onset date from all

150   individuals (Mean Midway and Mean Midway 7-Day). For all cases I noted the rate at

151   which this could not be calculated due to inadequate sampling (ie. if no individuals have

152   an absence observation within 7 days prior to the first presence, than no estimate can be

153   made).

154   The Weibull method fits a Weibull distribution to only the flowering presence

155   observations, thus is advantageous when no absence observations are available. The

156   flexible Weibull distribution can model a variety of shapes, and is commonly used to

157   used to estimate the start or end of a process. The estimated date of first flowering is the

158   sum of the dates of all flowering weighted by the joint Weibull distribution and is

159   equivalent to estimating an extinction date (Roberts and Solow 2003, Pearse et al. 2017).

160   This was used for both population and individual level estimates. Code for this in the R

161   language was obtained from Pearse et al. (2017) and is provided in the code repository.

162   The Logistic method fits a generalized linear model to both presence and absence

observations using a binomial distribution, where the DOY was used to explain the

presence or absence of flowering (glm(flowering ~ doy, family=binomial)). Prior to

fitting all flowering absence observations after the last observed flowering presence

were excluded. The expected probability of observing a flower was calculated for all

DOYs 1-365, and the estimated onset of flowering was the first DOY in the season in

which the expected probability exceeded a given threshold. The inverse of this is used to

estimate the end of flowering. All absence observations prior to the First Observed

flowering date were excluded, the expected probability was calculated for all DOYs

1-365, and the first DOY where the probability of flowering falls below the threshold

was the estimate for the end of flowering. I evaluated a range of probability thresholds

(0.05, 0.25, 0.50 ,0.75, and 0.95) and used the one with the highest R^2 for each

combination of metric, sample size, and flowering presence ratio. This method was used

in both the population and individual level analysis.

The GAM method is unique in that it can potentially estimate the full flowering

phenology for a season (onset, peak, and end) using smoothing splines. Similar to the

Logistic method, a general additive model was fit with a binomial distribution and DOY

explaining the presence or absence of flowers, where the DOY was a thin plate

regression spline (gam(flowering ~ s(doy, bs='tp'), family=binomial)). The expected

probability of flowering was calculated for all DOYs 1-365. The estimated onset date

was the first DOY in which the probability exceeded a given threshold. The estimated

peak flowering date was the DOY with the maximum probability in a given year. The

estimated end of flowering was the first DOY, after the peak, in which the probability

fell below the threshold. As in the Logistic method I evaluated five probability

thresholds and chose the one with the highest R^2 for each metric and scenario. Results

showing the best probability thresholds for the GAM and Logistic are available in

Figure S4. The GAM method was used for estimating onset and end in both the

population and individual level analysis, and for estimating peak flowering in the

population analysis.

The Survival method uses a Kaplan–Meier model, which is commonly used to estimate the survival of medical patients. Patient survival (alive or dead) observed in the years following a treatment is used in the model to estimate overall survival probability, with median survival rate, in years, used as a common summary statistic. In a phenology context observations of non-flowering and flowering can be ascribed to alive or dead, respectively, and the DOY, instead of year, of observation used as the time (Templ et al. 2017). The median survival rate can then be interpreted as the median time for flowering. I used the survfit function in the R package survival using right censoring (Therneau 2015). This method was used to estimate peak flowering in the population analysis.

Finally, the Mean Flowering method uses the average DOY of all flowering presence observations from throughout the year. This was used to estimate peak flowering in the population analysis.

All analysis was done using the R programming language (version 3.6.0, R Core Team 2017). Packages used during the analysis included dplyr (version 0.8.1, Wickham et al. 2017), tidyr (version 0.8.3, Wickham and Henry 2018), ggplot2 (version 3.1.1, Wickham 2016), mgcv (Wood 2003, version 1.8.28, 2011), survival (version 2.44.1.1, Therneau 2015), testthat (version 2.1.1, Wickham 2011), ggridges (version 0.5.1, Wilke 2018), and lubridate (version 1.7.4, Grolemund and Wickham 2011). Code to fully reproduce this analysis is available on GitHub (https://github.com/sdtaylor/phenology_estimators) and archived on Zenodo (https://doi.org/10.5281/zenodo.3234913).

Table 1: Estimators used in this analysis. p indicates the estimator uses only presence observations as opposed to both presence and absence observations.

	Population			Individual	
	Onset	Peak	End	Onset	End
First/Last Observed	✓ p		✓ p	✓ p	✓ p
Midway / Midway 7-Day				✓	✓
Mean Midway / Mean Midway 7-Day	✓		✓		
Weibull	✓ p		✓ p	✓ p	✓ p
Logistic	✓		✓	✓	✓
GAM	✓	✓	✓	✓	✓
Survival		✓			
Mean Flowering		✓ p			

Results

Population Onset Estimates

For population level flowering onset the Weibull method produced estimates with the lowest error for most scenarios (Fig. 1). Excluding the scenario where the proportion of flowering presence was 25% and with a sample size of 10, the Weibull method had R^2 values from 0.34 - 0.79 and median error rates of 3 - 4 days (upper and lower bound errors range from -20 - -1 and 15 - 8 for the 0.025 and 97.5 quantiles, respectively).

220 With a flowering proportion presence of 25% and sample size of 10 the First Observed

221 method had the highest R^2, but still overestimated the true dates by 11 days on average.

222 With higher sample sizes the First Observed method performed comparable to, but

223 always slightly worse than, the Weibull method.

224 The Logistic and GAM methods had the highest R^2, and similar median errors to the

225 Weibull method, when the sample size was high (50-100) and ratio of flowering

226 presence low (25%). In the scenarios where they had the highest R^2, the best threshold

227 for estimating onset was 0.25 and 0.50 for the GAM and Logistic methods, respectively

228 (Fig. S4). As the proportion of flowering presence increased, and relative amount of

229 absences decreased, the Logistic and GAM methods tended to perform worse (Fig. 1).

230 This was due to larger time gaps in the data since flowering presence observations occur

231 during a short time window. The gaps resulted in overfit models which increasingly

232 underestimated flowering onset as the proportion of flowering absences decreased (Fig

233 S5).

234 The Mean Midway and Mean Midway 7-Day methods were never the best performing

235 methods for estimating population onset. The Mean Midway method did not improve

236 by increasing the sample size or by increasing the proportion of flowering presence

237 observations. Results from the Mean Midway 7-Day method using a sample size of 10

238 were excluded due to less than 1% of random samples resulting in a usable estimate.

239 This was due to the requirement of each individual plant having at least one presence

240 and one prior absence observation. The usable number of estimates for the remaining

241 scenarios ranged from 2-10% (Fig. S1). With a sample size of 10 the GAM method

242 only produced estimates 27-81% of the time because of too few absence observations,

243 and 100% of the time in all other scenarios.
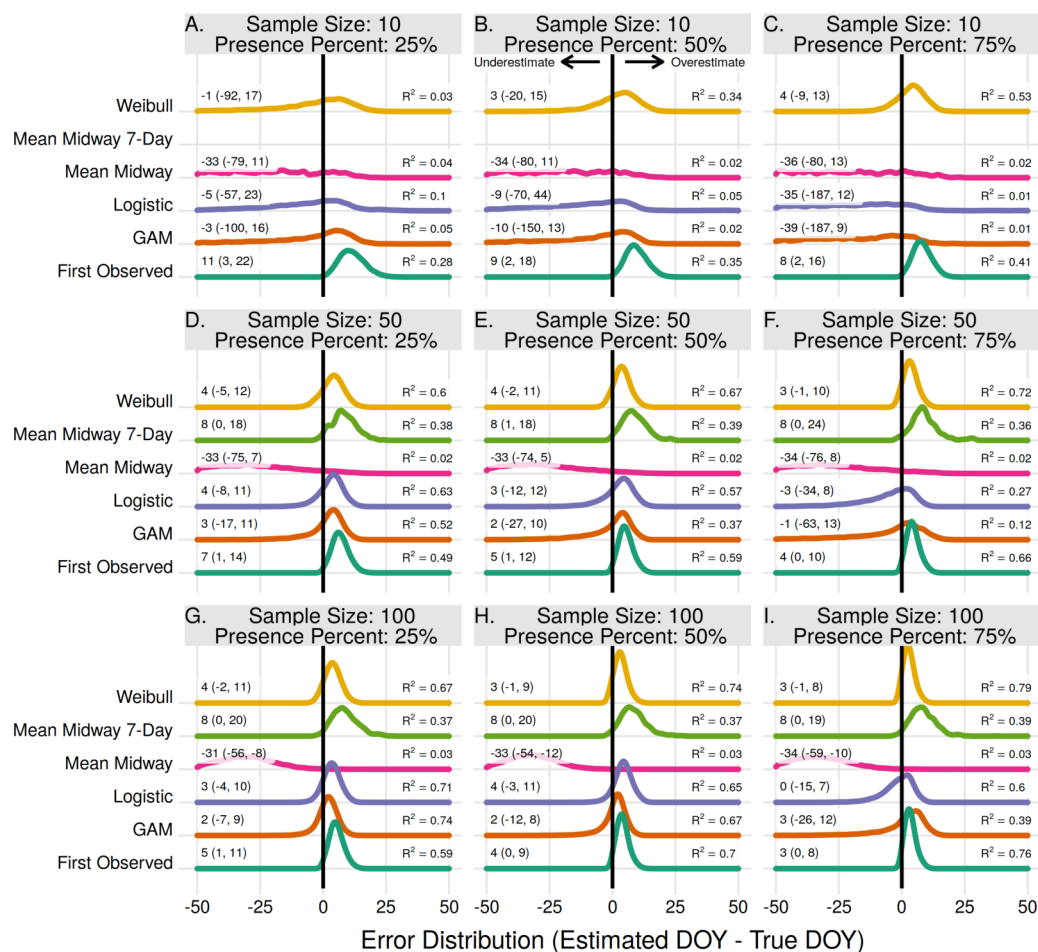
Figure 1: The error distribution of all estimators for population onset. The density curves are each derived from 11,000 randomly drawn observations, with sample size and presence proportion stated, across eleven years of phenological data. Text values represent the median error and the 95% quantile range in parenthesis.

Population End Estimates

The end of flowering for the entire population was more difficult to estimate than the onset of flowering. The highest R^2 for a given scenario in estimating population onset was always higher than the same scenario in population end estimates. For end estimates the Weibull method had the highest R^2 in 4 of 9 scenarios, including all three scenarios where the proportion of flowering presence was 75%, as well as when the proportion was 50% with a sample size of 10 (Fig. 2, B,C,F,I). With a sample size of 50 and 100 and a presence proportion of 50% and 25% the Logistic and GAM methods had the highest R^2 (Fig. 2, D,E,G,H). Where it performed the best the Logistic method used a threshold of 0.25 or 0.50 for estimating flowering end, while the GAM method used a threshold of 0.05 (Fig. S4). As in estimating population onset, the Logistic and GAM methods performed worse with increasing flowering presence due to large gaps in the absence data (Fig. S5).

With a sample size of 10 and presence proportion of 25% the Last Observed method had the highest R^2, but still underestimated the end date of flowering by 17 days the majority of the time (Fig. 2,A). The Midway method, both with and without the 7-day restriction, were never the best performing estimators. Without the 7-day restriction the method consistently overestimated the end date. With the 7-day restriction the method consistently underestimated the end date. Neither Midway method improved with either increasing sample size or increasing proportion of flowering presence. As in the population onset the results from the Mean Midway 7-Day method were excluded due to less than 1% of estimates being usable, and the GAM method had a low proportion (27-81%) of usable estimates with a sample size of 10 (Fig. S1).
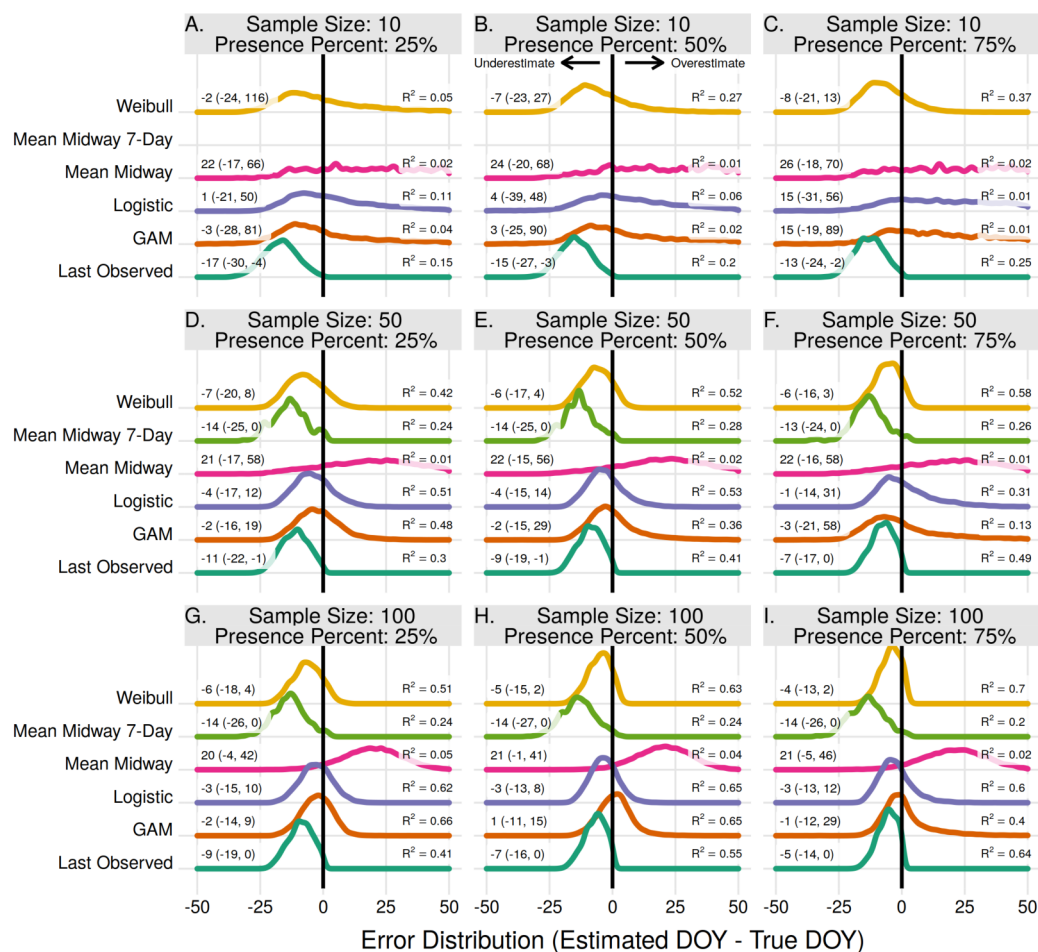
Figure 2: The error distribution of all estimators for population end. The density curves are each derived from 11,000 randomly drawn observations, with sample size and presence proportion stated, across eleven years of phenological data. Text values represent the median error and the 95% quantile range in parenthesis.

Population Peak Estimates

All three methods to estimate peak flowering had median error rates of 1 day except in one instance, using the GAM method for a sample size of 10 and proportion of flower presence 75% (Fig. 3, C). The Mean Flowering method had the highest R^2 in all scenarios except three where it had R^2 values equal to the Survival Curve method. For the Mean and Survival Curve methods, errors improved with both increasing sample size and increasing proportion of flowering presence. For the GAM method errors improved with increasing sample size, but worsened with increasing proportion of flowering presence.

Individual Onset and End Estimates

For individual plant onset estimates the Midway 7-Day method performed the best in 7 of 9 scenarios (Fig. 4, A-C,D,E,G,H). In two scenarios, when the sample size was 15 and 20 with a proportion of flowering presence observations of 75%, the First Observed method had slightly higher R^2 and lower median error rates than the Midway 7-Day method (Fig. 4, F,I). The Midway 7-Day method was able to produce usable individual estimates only 3-17% of the time due to its restrictive nature, while the Midway method produced usable estimates 70-100% of the time (Fig. S2).

The Weibull, Midway, Logistic, and GAM methods never produced the best estimate for any scenario in estimating individual plant flowering onset. The Weibull method did improve with increasing sample size and increasing proportion of flowering presence. Though, since the Weibull method does not use absence observations, increasing the flowering proportion effectively just increases the sample size. At effective sample sizes of 10 or more the Weibull method produced estimates only slightly worse than the Midway 7-Day and First Observed method. The Midway, Logistic, and GAM methods improved slightly with increasing sample size but worsened with increasing proportion
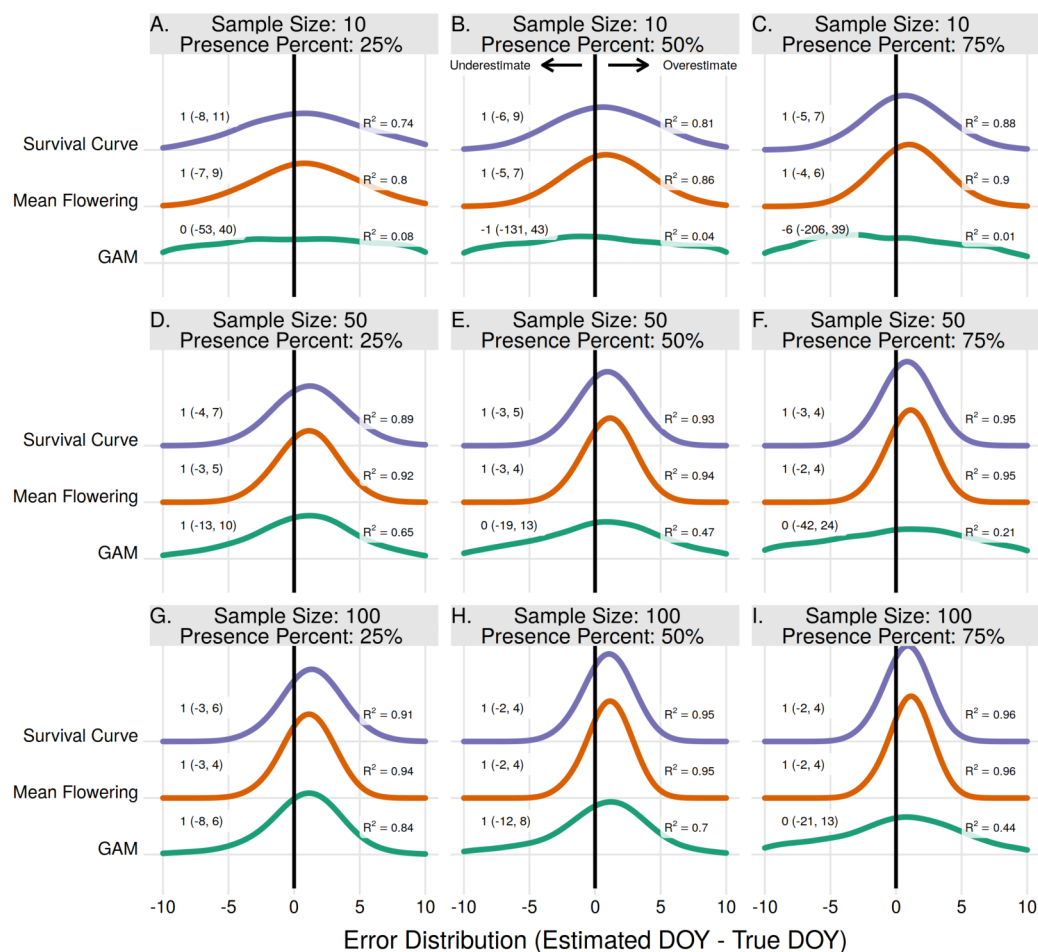
Figure 3: The error distribution of all estimators for population peak. The density curves are each derived from 11,000 randomly drawn observations, with sample size and presence proportion stated, across eleven years of phenological data. Text values represent the median error and the 95% quantile range in parenthesis.
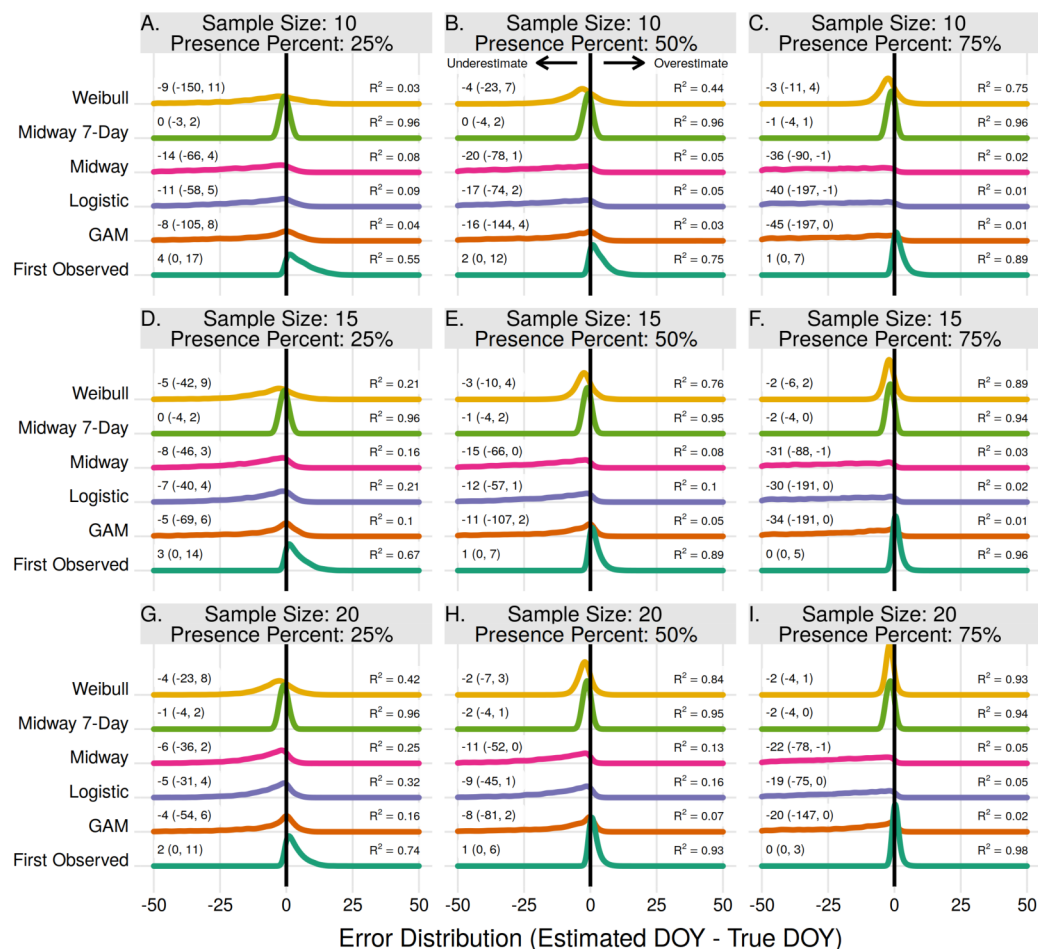
Figure 4: The error distribution of all estimators for individual onset. The density curves are each derived from 4840 randomly drawn observations, with sample size and presence proportion stated, across eleven years of phenological data for 286 individual plants. Text values represent the median error and the 95% quantile range in parenthesis.

292 of flowering presence.

293 The errors from individual end estimates were nearly identical to individual onset errors,

294 thus the model performance outcomes were the same. Individual end errors are supplied

295 in the supplement (Fig. S3).

Discussion

Overall findings

This comparison of phenological estimators using a dataset with known onset, peak, and end of flowering dates confirmed biases in some estimators and shows the strength of newer ones. Overall the Weibull method predominantly outperformed all other methods for estimating the onset and end of flowering populations. The Mean Flowering method produced better, or equal, estimates than other methods for flowering peak. The Midway 7-Day method outperformed other methods in estimating onset and end of individuals flowering, albeit with limitations on the usable sample size. Exceptions to these stem mainly from differences in sample size but also the shape of the flowering distribution.

The Weibull method was the best overall for estimating population onset and end with two exceptions. First, when the total number of flowering presence observations were extremely low (ie. with a total sample size of 10 and percent presence observations 25%) using just the first or last observed flowering date produced better estimates. Yet with such a low R^2 values this method cannot be recommended, and along with other studies I recommend not estimating flowering onset or end with extremely low sample sizes (Miller-Rushing et al. 2008, Moussus et al. 2010, Bertin 2015). Second, using a larger sample size (50-100) and a small proportion of flowering presence the Logistic and GAM methods performed slightly better than the Weibull method. This suggests the Logistic and GAM methods effectively utilize flowering absence observations, but require a large amount of them, relative to presence observations, to accurately describe the phenology. Exploring the GAM and Logistic methods further showed that regular sampling, especially during the non-flowering season may also be important. Absence observations are rare in herbarium data due to a bias toward growing season sampling (Rich and Woodruff 1992, Daru et al. 2018), but more common in datasets with status-based protocols (Denny et al. 2014, Elmendorf et al. 2016). Given that flowering

322 absence observations could prove useful when presence observations are low, absence

323 observations should be emphasized in future data collection efforts.

324 For estimating the peak of flowering populations the Mean Flowering method

325 consistently produced the best estimate, even when the sample size and proportion of

326 flowering presence was low, 10 and 25%, respectively. This method has the advantage

327 over the Survival Curve and GAM method of not requiring flowering absence

328 observations. As noted in other studies the Mean Flowering method is a reliable method

329 for estimating peak flowering (Miller-Rushing et al. 2008, Moussus et al. 2010, Bertin

330 2015).

331 For estimating the start and end of flowering for individual plants the Midway 7-Day

332 method was the best in most cases. The Weibull method performed similarly when the

333 absolute number of flowering presence observations was greater than 10, and the First

334 Observed method also performed well with a high amount of presence observations.

335 The First Observed method can be advantageous as it ensures no underestimate of the

336 onset date (or no overestimate of the end date if using Last Observed). In the vast

337 majority of cases (83-97% depending on the scenario, Fig. S2) it was not possible to use

338 the Midway 7-Day method due to lack of individuals with an absence observation

339 within 7 days of the first presence observation. With large enough datasets using this

340 method is still possible even with the restriction (Gerst et al. 2016), and it can also be

341 relaxed with a 15 or 30 day minimum to increase sample size if needed (Taylor et al.

342 2019). If an insufficient number of individuals results from applying the restriction

343 using the Midway method, then the Weibull or First Observed methods are preferable

344 for estimating onset in an individual given enough flowering presence observations.

345 While the Midway 7-Day method was the best for estimating flowering for individual

346 plants, using the mean of those estimates from a population (Mean Midway 7-Day) did

347 not provide the best population level estimates even with a large sample size. The onset

348 of flowering for individuals is staggered over time and the mean of these start times is

not equivalent to the population onset date (Ison and Wagenius 2014, Keyzer et al. 2017,

Renzi et al. 2019).

Prior study comparison

Moussus et al. (2010) found GAM's to be among the best estimators for detecting

phenological shifts among different seasons, yet here the GAM method performed best

only in scenarios with a large proportion of flowering absences. Differences in analysis

include Moussus et al. (2010) using a poisson distribution with simulated count data,

while here I used a binomial distribution and observed presence/absence data. Moussus

et al. (2010) also did not evaluate the Weibull estimator, which outperformed the GAM

method in many scenerios in the current study. Here the performance of the GAM

method was influenced by the proportion of absence observations, where their relative

amount affected the best threshold to use as well as the highest accuracy attainable (Fig.

S5). Future studies could potentially adjust the GAM model specifications to better

accommodate scenarios with a low proportion of absence observations. It is also

possible that the output for the GAM model used here, the probability of observing a

flower, is not analogous to the total abundance of flowers. General additive models have

substantial flexibility (Wood 2017, Simpson 2018, Pedersen et al. 2018) and further

exploration into their use for plant phenology would be beneficial.

Drivers of estimator performance

The shape of the flowering distribution affected the proficiency of the estimators. The

number of *E. angustifolia* flowers observed over time resembles a skewed distribution,

with a quick onset, peak, and gradual decline in number of flowers. The long tail made

end estimates more difficult as the probability of observing a flower close to the true end

was low. The best performing estimators also tended to overestimate onset and

373 underestimate end of population flowering, as the majority of randomly sampled

374 observations came from the center of the flowering period. The likeness to a normal

375 distribution allowed for very accurate estimates of peak flowering using the Mean

376 Flowering method. Flowering distributions for many species are thought to have similar

377 properties (Forrest and Miller-Rushing 2010, Clark and Thompson 2011), but the

378 methods used here may not be appropriate for other phenophases, especially ones which

379 can last significantly longer (ie. leaves lasting several months on the tree). Flowering is

380 also expected to have non-uniform shifts from changing drivers (Ogilvie and Forrest

381 2017, Theobald et al. 2017). Other phenophases which do not have a distinct transition

382 or cannot be easily modelled using presence and absence, such as fruit maturation, may

383 not be well described by the methods used here. In these cases models integrating the

384 continuous cycle of phenology would likely need to be developed, such as using

385 integrated process based models (Chuine and Régnière 2017) or hierarchical bayesian

386 models (Clark et al. 2014).

387 The outcomes for estimating the end of individuals flowering was essentially identical

388 to estimates for the onset. The flowering of an individual *E. angustifolia* plant over time

389 approximates a uniform distribution. Thus, unlike the skewed population flowering over

390 time, estimators for the onset and end of individuals perform equally. This may not be

391 the case when the study species are larger in size and/or contain numerous flowers

392 which can be counted (Renzi et al. 2019). In these cases the phenology over time may

393 be more similar to a population, with a flowering peak and potentially skewed

394 distribution (while *E. angustifolia*, being in the family Asteraceae, can have one or more

395 flowering heads each with numerous florets, here I treated each individual plant as a

396 single unit).

Recommendations

Results from this study can be applied to two common sources of large-scale status-based phenological observations, herbarium data and citizen science data. Data from herbarium specimens represent spatially diffuse observations at the population scale or larger, with a bias toward flowering presence (Willis et al. 2017, Daru et al. (2018)). The best onset estimator for these data depends on the type and amount of data available. With a low sample size (less than 10 observations) I recommend not estimating onset as it can lead to high errors (Miller-Rushing et al. 2008, Moussus et al. 2010, Bertin 2015). With larger sample sizes the Weibull method will be appropriate in most cases as herbarium data are mostly presence observations, but when there are a large amount of absences the GAM or Logistic methods should be explored. With a very large sample size (greater than 50) the First Observed method can be just as accurate as the Weibull, but note that this accuracy will likely decrease for longer lasting phenophases such as leaves or fruit. For estimating the end of a phenophase the same recommendations apply, with the caveat that the minimum sample size will need to increase if the phenophase distribution has a long tail. As herbarium specimens do not represent repeated observations of the same individual, individual level estimates are not applicable.

Citizen science phenological data can be subset into two types: 1) those from social media applications using geotagged images (ie. Twitter or iNaturalist, Silva et al. 2018), and 2) those from observing networks and consisting of repeated observations of the same site or individual plant (ie. the USA National Phenology Network or Pan European Phenological database, Denny et al. 2014, Templ et al. 2018). For the former the same recommendations as from herbarium specimen data apply. For the latter, if estimates for individual plants are needed then the Midway-7 Day method is most suitable as long as absence observations are available and the final usable amount of data is adequate. Without absences, or to provide more usable data, the First Observed
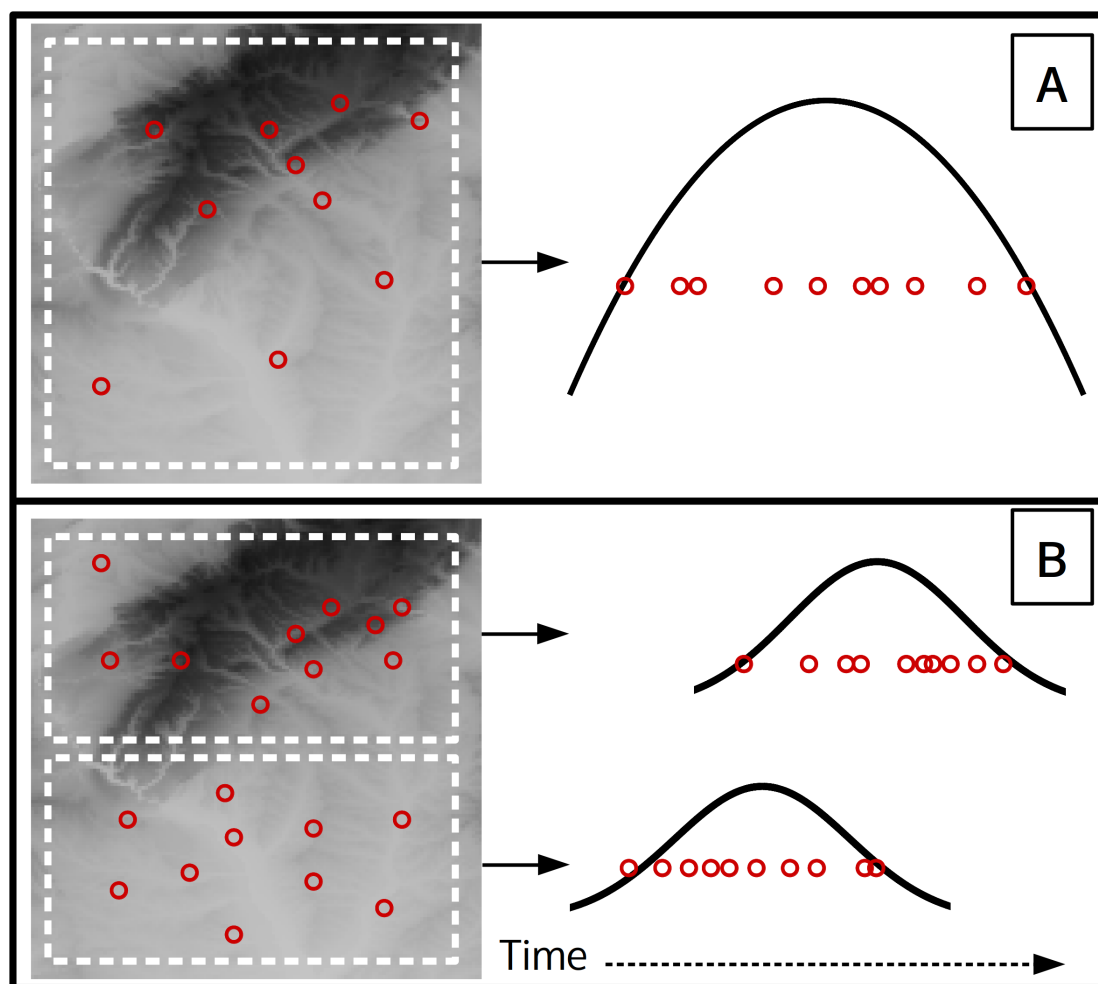
Figure 5: A theoretical scenario where 10 flowering observations are used to estimate onset across a landscape (A), and a second scenario where onset is estimated at a finer spatial grain on the same landscape (B). Curves indicate the theoretical flowering distribution over time for the respective area.

424 method can be used as long as the sample size is adequate, and the Weibull method

425 should be considered regardless due to its ability to generate confidence intervals

426 (Pearse et al. 2017).

427 This analysis used data from a single site, yet with herbarium or citizen science data

428 observations more commonly represent a large spatial extent. At these larger scales the

429 underlying phenology of a species is affected by an array of biotic and abiotic factors

430 which can cause different flowering times at distant locations (Diez et al. 2012, Keyzer

431 et al. 2017, Prevéy et al. 2017). When combining phenological observations from

⁴³² different locations any transition estimates will be for some subset of the full flowering

⁴³³ phenology across the species entire range (ie. the universal curve, Keyzer et al. (2017)).

⁴³⁴ The spatial extent and grain of the analysis will affect the minimum sample size needed

⁴³⁵ and also what the estimates represent due to the modifiable areal unit problem (Jelinski

⁴³⁶ and Wu 1996). For example consider a case where 10 observations of flowering from a

⁴³⁷ single year are used to estimate onset (Fig. 5A), which represents flowering onset for

⁴³⁸ the entire landscape. If the same landscape is subset to a finer spatial grain (Fig. 5B),

⁴³⁹ then each of the two smaller spatial units could have an independent onset estimate, but

⁴⁴⁰ would each require an adequate sample size. Also note that the onset estimate for the

⁴⁴¹ larger grain (Fig. 5A) will likely approximate the onset estimate of the earlier of the two

⁴⁴² smaller grain estimates, while the larger grain end estimate will approximate the later of

⁴⁴³ the smaller grain estimates. Previous studies used political boundaries as the spatial unit

⁴⁴⁴ (Park 2014, Pearse et al. 2017), though the optimal spatial grain and observation density

⁴⁴⁵ needed likely depends on the species being analysed and the large-scale gradients over

⁴⁴⁶ which it occurs. Future studies should examine these relationships between spatial

⁴⁴⁷ scales and phenology more closely.

⁴⁴⁸ Conclusion

⁴⁴⁹ In summary I have used a precise flowering phenological dataset to confirm that naively

⁴⁵⁰ using the first flowering observation is biased, and estimates using the Mean Flowering

⁴⁵¹ reliable for estimating flowering peak. I have also shown how the recently introduced

⁴⁵² Weibull method can produce reliable estimates given an adequate sample size. The

⁴⁵³ Logistic and GAM methods can be useful with large datasets having low amounts of

⁴⁵⁴ flowering presence, and future collection efforts should emphasize absence observations

⁴⁵⁵ for this reason. Additionally, estimating transition dates of individual plants is best done

⁴⁵⁶ with the Midway method using a 7 day restriction, and the Weibull method if the

⁴⁵⁷ restriction results in a low number of final samples. These estimators are needed for

458 translating status-based phenological data into distinct transition dates used to track

459 changing seasonal patterns.

Acknowledgements

461 I thank Amy Waananen and the team of the Echinacea Project

462 (http://echinaceaproject.org/) for providing the phenology data set. I thank Amy

463 Waananen and Janet Prevéy for providing feedback on early versions of the manuscript.

References

465 Basler, D. 2016. Evaluating phenological models for the prediction of leaf-out dates in

466 six temperate tree species across central europe. Agricultural and Forest Meteorology

467 217:10–21.

468 Bertin, R. I. 2015. Climate change and flowering phenology in worcester county,

469 massachusetts. International Journal of Plant Sciences 176:107–119.

470 Calinger, K. M., S. Queenborough, and P. S. Curtis. 2013. Herbarium specimens reveal

471 the footprint of climate change on flowering trends across north-central north america.

472 Ecology Letters 16:1037–1044.

473 CaraDonna, P. J., A. M. Iler, and D. W. Inouye. 2014. Shifts in flowering phenology

474 reshape a subalpine plant community. Proceedings of the National Academy of

475 Sciences 111:4916–4921.

476 Chuine, I., and J. Régnière. 2017. Process-based models of phenology for plants and

477 animals. Annual Review of Ecology, Evolution, and Systematics 48:159–182.

478 Clark, J. S., J. Melillo, J. Mohan, and C. Salk. 2014. The seasonal timing of warming

that controls onset of the growing season. Global Change Biology 20:1136–1145.

Clark, R. M., and R. Thompson. 2011. Estimation and comparison of flowering curves. Plant Ecology & Diversity 4:189–200.

Daru, B. H., D. S. Park, R. B. Primack, C. G. Willis, D. S. Barrington, T. J. S. Whitfeld, T. G. Seidler, P. W. Sweeney, D. R. Foster, A. M. Ellison, and C. C. Davis. 2018. Widespread sampling biases in herbaria revealed from large-scale digitization. New Phytologist 217:939–955.

Davis, C. C., C. G. Willis, B. Connolly, C. Kelly, and A. M. Ellison. 2015. Herbarium records are reliable sources of phenological change driven by climate and provide novel insights into species' phenological cueing mechanisms. American Journal of Botany 102:1599–1609.

Denny, E. G., 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, and J. F. Weltzin. 2014. Standardized phenology monitoring methods to track plant and animal activity for science and resource management applications. International Journal of Biometeorology 58:591–601.

Dickinson, J., B. Zuckerberg, and D. Bonter. 2010. Citizen science as an ecological research tool: Challenges and benefits. Annual Review of Ecology, Evolution and Systematics 41:149–172.

Diez, J. M., I. Ibáñez, A. J. Miller-Rushing, S. J. Mazer, T. M. Crimmins, M. A. Crimmins, C. D. Bertelsen, and D. W. Inouye. 2012. Forecasting phenology: From species variability to community patterns. Ecology Letters 15:545–553.

Elmendorf, S. C., K. D. Jones, B. I. Cook, J. M. Diez, C. A. F. Enquist, R. A. Hufft, M. O. Jones, S. J. Mazer, A. J. Miller-Rushing, D. J. P. Moore, M. D. Schwartz, and J. F. Weltzin. 2016. The plant phenology monitoring design for the national ecological

504   observatory network. Ecosphere 7:e01303.

505   Forrest, J., and A. J. Miller-Rushing. 2010. Toward a synthetic understanding of the

506   role of phenology in ecology and evolution. Philosophical Transactions of the Royal

507   Society B: Biological Sciences 365:3101–3112.

508   Gerst, K. L., J. L. Kellermann, C. A. F. Enquist, A. H. Rosemartin, and E. G. Denny.

509   2016. Estimating the onset of spring from a complex phenology database: Trade-offs

510   across geographic scales. International Journal of Biometeorology 60:391–400.

511   Grolemund, G., and H. Wickham. 2011. Dates and times made easy with {lubridate}.

512   Journal of Statistical Software 40:1–25.

513   Ison, J. L., and S. Wagenius. 2014. Both flowering time and distance to conspecific

514   plants affect reproduction in echinacea angustifolia , a common prairie perennial.

515   Journal of Ecology 102:920–929.

516   Jelinski, D. E., and J. Wu. 1996. The modifiable areal unit problem and implications for

517   landscape ecology. Landscape Ecology 11:129–140.

518   Jones, C. A., and C. C. Daehler. 2018. Herbarium specimens can reveal impacts of

519   climate change on plant phenology; a review of methods and applications. PeerJ

520   6:e4576.

521   Keyzer, C. W. de, N. E. Rafferty, D. W. Inouye, and J. D. Thomson. 2017. Confounding

522   effects of spatial variation on shifts in phenology. Global Change Biology

523   23:1783–1791.

524   Lindén, A., K. Meller, and J. Knape. 2017. An empirical comparison of models for the

525   phenology of bird migration. Journal of Avian Biology 48:255–265.

526   Linkosalo, T., H. K. Lappalainen, and P. Hari. 2008. A comparison of phenological

527   models of leaf bud burst and flowering of boreal trees using independent observations.

528   Tree physiology 28:1873–1882.

529   Mayor, S. J., R. P. Guralnick, M. W. Tingley, J. Otegui, J. C. Withey, S. C. Elmendorf,

530   M. E. Andrew, S. Leyk, I. S. Pearse, and D. C. Schneider. 2017. Increasing

531   phenological asynchrony between spring green-up and arrival of migratory birds.

532   Scientific Reports 7:1902.

533   Miller-Rushing, A. J., D. W. Inouye, and R. B. Primack. 2008. How well do first

534   flowering dates measure plant responses to climate change? The effects of population

535   size and sampling frequency. Journal of Ecology 96:1289–1296.

536   Moussus, J.-P., R. Julliard, and F. Jiguet. 2010. Featuring 10 phenological estimators

537   using simulated data. Methods in Ecology and Evolution 1:140–150.

538   Moussus, J., F. Jiguet, J. Clavel, and R. Julliard. 2009. A method to estimate

539   phenological variation using data from large-scale abundance monitoring programmes.

540   Bird Study 56:198–212.

541   Newson, S. E., N. J. Moran, A. J. Musgrove, J. W. Pearce-Higgins, S. Gillings, P. W.

542   Atkinson, R. Miller, M. J. Grantham, and S. R. Baillie. 2016. Long-term changes in the

543   migration phenology of uk breeding birds detected by large-scale citizen science

544   recording schemes. Ibis 158:481–495.

545   Ogilvie, J. E., and J. R. Forrest. 2017. Interactions between bee foraging and floral

546   resource phenology shape bee populations and communities. Current Opinion in Insect

547   Science 21:75–82.

548   Park, I. W. 2014. Impacts of differing community composition on flowering phenology

549   throughout warm temperate, cool temperate and xeric environments. Global Ecology

550   and Biogeography 23:789–801.

551   Pearse, W. D., C. C. Davis, D. W. Inouye, R. B. Primack, and T. J. Davies. 2017. A

552   statistical estimator for determining the limits of contemporary and historic phenology.

553   Nature Ecology & Evolution 1:1876–1882.

554   Pedersen, E. J., D. L. Miller, G. L. Simpson, and N. Ross. 2018. Hierarchical

555   generalized additive models: An introduction with mgcv. PeerJ Preprints 6:e27320v1.

556   Prevéy, J., M. Vellend, N. Rüger, R. D. Hollister, A. D. Bjorkman, I. H. Myers-Smith, S.

557   C. Elmendorf, K. Clark, E. J. Cooper, B. Elberling, A. M. Fosaa, G. H. R. Henry, T. T.

558   Høye, I. S. Jónsdóttir, K. Klanderud, E. Lévesque, M. Mauritz, U. Molau, S. M. Natali,

559   S. F. Oberbauer, Z. A. Panchen, E. Post, S. B. Rumpf, N. M. Schmidt, E. A. G. Schuur,

560   P. R. Semenchuk, T. Troxler, J. M. Welker, and C. Rixen. 2017. Greater temperature

561   sensitivity of plant phenology at colder sites: Implications for convergence across

562   northern latitudes. Global Change Biology 23:2660–2671.

563   R Core Team. 2017. R: A language and environment for statistical computing. Vienna:

564   R Foundation for Statistical Computing.

565   Renzi, J. J., W. D. Peachey, and K. L. Gerst. 2019. A decade of flowering phenology of

566   the keystone saguaro cactus (Carnegiea gigantea). American Journal of Botany

567   106:199–210.

568   Rich, T., and E. Woodruff. 1992. Recording bias in botanical surveys. Watsonia:73–95.

569   Roberts, D. L., and A. R. Solow. 2003. When did the dodo become extinct? Nature

570   426:245–245.

571   Schaber, J., and F.-W. Badeck. 2002. Evaluation of methods for the combination of

572   phenological time series and outlier detection. Tree Physiology 22:973–982.

573   Scheffers, B. R., L. De Meester, T. C. L. Bridge, A. A. Hoffmann, J. M. Pandolfi, R. T.

574   Corlett, S. H. M. Butchart, P. Pearce-Kelly, K. M. Kovacs, D. Dudgeon, M. Pacifici, C.

575   Rondinini, W. B. Foden, T. G. Martin, C. Mora, D. Bickford, and J. E. M. Watson. 2016.

576   The broad footprint of climate change from genes to biomes to people. Science

577 354:aaf7671.

578 Silva, S. J., L. K. Barbieri, and A. K. Thomer. 2018. Observing vegetation phenology

579 through social media. PLOS ONE 13:e0197325.

580 Simpson, G. L. 2018. Modelling palaeoecological time series using generalized additive

581 models. bioRxiv:322248.

582 Taylor, S. D., J. M. Meiners, K. Riemer, M. C. Orr, and E. P. White. 2019. Comparison

583 of large-scale citizen science data and long-term study data for phenology modeling.

584 Ecology 100:e02568.

585 Templ, B., S. Fleck, and M. Templ. 2017. Change of plant phenophases explained by

586 survival modeling. International Journal of Biometeorology 61:881–889.

587 Templ, B., E. Koch, K. Bolmgren, M. Ungersböck, A. Paul, H. Scheifinger, T.

588 Rutishauser, M. Busto, F.-M. Chmielewski, L. Hájková, S. Hodzić, F. Kaspar, B.

589 Pietragalla, R. Romero-Fresneda, A. Tolvanen, V. Vučetič, K. Zimmermann, and A.

590 Zust. 2018. Pan European Phenological database (PEP725): a single point of access for

591 European data. International Journal of Biometeorology 62:1109–1113.

592 Theobald, E. J., I. Breckheimer, and J. HilleRisLambers. 2017. Climate drives

593 phenological reassembly of a mountain wildflower meadow community. Ecology

594 98:2799–2812.

595 Therneau, T. M. 2015. A package for survival analysis in s. version 2.43.3.

596 Waananen, A., G. Kiefer, J. L. Ison, and S. Wagenius. 2018a. Data from: Mating

597 opportunity increases with synchrony of flowering among years more than synchrony

598 within years in a nonmasting perennial. Dryad Digital Repository.

599 Waananen, A., G. Kiefer, J. L. Ison, and S. Wagenius. 2018b. Mating opportunity

600 increases with synchrony of flowering among years more than synchrony within years

in a nonmasting perennial. The American Naturalist 192:379–388.

Wagenius, S. 2004. Style persistence, pollen limitation, and seed set in the common prairie plant echinacea angustifolia (asteraceae). International Journal of Plant Sciences 165:595–603.

Wickham, H. 2011. testthat: Get started with testing. The R Journal 3:5–10.

Wickham, H. 2016. ggplot2: Elegant graphics for data analysis. Springer-Verlag New York.

Wickham, H., and L. Henry. 2018. tidyr: Easily tidy data with 'spread()' and 'gather()' functions. R package version 0.8.3.

Wickham, H., R. Francois, L. Henry, and K. Müller. 2017. dplyr: A grammar of data manipulation. R package version 0.8.1.

Wilke, C. O. 2018. ggridges: Ridgeline plots in 'ggplot2'. R package version 0.5.1.

Willis, C. G., E. R. Ellwood, R. B. Primack, C. C. Davis, K. D. Pearson, A. S. Gallinat, J. M. Yost, G. Nelson, S. J. Mazer, N. L. Rossington, T. H. Sparks, and P. S. Soltis. 2017. Old plants, new tricks: Phenological research using herbarium specimens. Trends in Ecology & Evolution 32:531–546.

Wolkovich, E. M., B. I. Cook, J. M. Allen, T. M. Crimmins, J. L. Betancourt, S. E. Travers, S. Pau, J. Regetz, T. J. Davies, N. J. B. Kraft, T. R. Ault, K. Bolmgren, S. J. Mazer, G. J. McCabe, B. J. McGill, C. Parmesan, N. Salamin, M. D. Schwartz, and E. E. Cleland. 2012. Warming experiments underpredict plant phenological responses to climate change. Nature 485:494–497.

Wood, S. N. 2003. Thin plate regression splines. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 65:95–114.

Wood, S. N. 2011. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. Journal of the Royal Statistical

626   Society: Series B (Statistical Methodology) 73:3–36.

627   Wood, S. N. 2017. Generalized additive models: An introduction with r. CRC Press,

628   Boca Raton, FL.

629   Zhang, H., S. Liu, P. Regnier, and W. Yuan. 2017. New insights on plant phenological

630   response to temperature revealed from long-term widespread observations in china.

631   Global Change Biology 12:3218–3221.