PhenoForecaster: A software package for the prediction of flowering phenology

Isaac Park, Alex Jones, and Susan J. Mazer

PREMISE OF THE STUDY: Predicting the flowering times of angiosperm taxa is a goal of mounting importance in the face of future climate change, with applications not only in plant biology and ecology, but also horticulture, agriculture, and invasive species management. To date, no tool is available to facilitate predictions of flowering phenology using multivariate phenoclimatic models. Such a tool is needed by researchers and other stakeholders who need to predict phenological activity, but are unfamiliar with phenoclimatic modeling techniques. PhenoForecaster allows users of any background to conduct species-specific phenological predictions using an intuitive graphical interface and provides an estimate of each prediction’s accuracy.

METHODS AND RESULTS: Elastic net regression techniques were used to develop species-specific models capable of predicting the flowering dates of 2320 angiosperm species.

CONCLUSIONS: PhenoForecaster is the first stand-alone package to make phenological modeling directly accessible to users without the need for in-depth phenological observations.

KEY WORDS: bloom timing; elastic net regularization; flowering; herbarium specimens; phenoclimate models; phenological models; phenology.
In this paper, we present PhenoForecaster, a software package that allows users to predict quickly and easily the mean flowering date for each of 2320 angiosperm species. PhenoForecaster uses readily accessible climate data in combination with species-specific phenological models that were generated by the authors using a simplified version of a method previously used to evaluate phenological responses to climate using digital herbarium records (Park and Mazer, 2018). Specifically, PhenoForecaster uses estimates of five climate parameters (i.e., the quantity of winter and spring precipitation that fell as snow in a given year [PAS_wt and PAS_sp, respectively], the number of frost-free days that occurred in a given winter and spring [NFFD_wt and NFFD_sp, respectively], and the date of the beginning of the frost-free period [BFFP expressed as the Day of Year]) to predict the day of year (DOY) on which the selected angiosperm species will reach its mean flowering date (MFD) at a location experiencing those conditions. These parameters represent the climate cues to which MFD was found to be most sensitive across the majority of these species using similar data and modeling techniques to those used by PhenoForecaster (Park and Mazer, 2018). In order to facilitate PhenoForecaster’s use, all of the phenoclimate models that it uses were limited to these climate parameters, which were sufficient to retain the majority of the predictive power produced by more complicated models (Park and Mazer, 2018). This package allows both manual entry of climate parameters as well as bulk entry of data in cases where phenological predictions are required across multiple locations or climate scenarios. PhenoForecaster has been designed to accept climate input in a comma-separated value (CSV) format that is compatible with climate data generated by ClimateNA (Wang et al., 2016), a freely available software package that produces spatially explicit estimates of historical climate conditions throughout North America, and which utilizes a user-friendly graphical interface and requires only that the user provide the latitude and longitude (either manually or as a CSV file) of all points of interest.

Thus, while predictions of phenological timing for a given plant species previously required extensive observation, modeling, and calculation, PhenoForecaster represents a simple-to-use tool through which the phenology of many angiosperm species can be readily predicted under any observed or theoretical climate scenario.

**METHODS AND RESULTS**

PhenoForecaster is an open source program written in Python 2.7 (Python Core Team, 2008). The source code, as well as the code used to construct the phenoclimate models used by PhenoForecaster, are available at [https://github.com/isaacWpark/PhenoForecaster](https://github.com/isaacWpark/PhenoForecaster). A Windows installer package and source code, as well as a user’s manual, are also available at [https://labs.eemb.ucsb.edu/mazer/susan/software](https://labs.eemb.ucsb.edu/mazer/susan/software). To install the package, the user simply needs to download and run the installer. The executable has been successfully tested on Windows 7, 8, and 10. PhenoForecaster has an intuitive graphical user interface that allows users with minimal prior experience with phenological prediction or with PhenoForecaster to predict the phenological timing of any targeted species by implementing the following steps.

First, the user must select the subset of species-specific models from which they wish to choose, based on the minimum model reliability they desire. By default, only the 490 species-specific models for which expected mean absolute error (MAE) ≤15 days (indicating that phenological predictions are, on average, typically within 15 days of observed MFDs) were considered to be “good” model fits, and are therefore displayed for selection. Depending on user preference, however, this list of species may be expanded to include species-specific models that exhibit higher MAE, or contracted to only display those species for which more accurate phenological models are available (Fig. 1A). Having filtered the species by the minimum MAE desired, the user must then use the species selection dropdown menu to select the species for which phenological predictions are to be generated (Fig. 1B). Second, the specific climatic conditions for which phenological predictions are desired may then be entered manually (Fig. 1C) or uploaded as a CSV data file (Fig. 1D). For the latter, the first line of the input file is a header line with column descriptions. The first two columns of the file, labeled 'ID1' and 'ID2', represent any string data the user desires to include for the purpose of identifying each row of data in a unique fashion. The remaining columns may be in any order, but must include the following: 'NFFD_wt', 'NFFD_sp', 'PAS_wt', 'PAS_sp', and 'BFFP'. Data in the column 'NFFD_wt' should consist of a count of the number of frost-free days from January 1 to March 31 in the year for which flowering time is to be estimated. Data in the column 'NFFD_sp' should consist of a count of the number of frost-free days from April 1 to June 30 in the year for which flowering time is to be estimated. Data in the column 'PAS_wt' should consist of the total precipitation that fell as snow (in mm) from April 1 to June 30 in the year in which flowering time is to be estimated. Data in the column 'BFFP' should consist of the DOY on which the annual frost-free period began. PhenoForecaster allows any number of additional data columns (representing other data that may be associated with each location or year to be predicted) to be placed into the input file. In cases where the user desires that data from such additional columns be preserved in the output file created by PhenoForecaster, they may select the ‘retain all input data’ option in the lower left of the user interface. If this option is selected, PhenoForecaster will preserve all columns from the input data, appending a new column with the header ‘DOY_Predicted’ that consists of the predicted MFD for a given row of data, and output all data as a CSV file. Otherwise, PhenoForecaster will generate output in the form of a CSV file, with the headers 'ID1', 'ID2', and 'DOY_Predicted'. PhenoForecaster utilizes phenoclimate models that were constructed for each species from herbarium-based phenological data using a total of 556,322 digital records of herbarium specimens collected in flower across 72 herbaria throughout North America (see Acknowledgments for complete listing), collected between 1901 and 2015 and structured in Darwin Core format. Specimens that did not include either the decimal latitude and longitude from which the sample was collected or the precise date of collection were eliminated. Specimens that were not explicitly recorded as being in flower within either the Darwin Core fields ‘reproductive-condition’ or ‘lifestage’ were eliminated. Specimens that were only listed as ‘in bud’ or ‘fructing’ were not considered to be in flower for purposes of this analysis. Duplicate specimens (i.e., specimens of a given species that were collected on the same date and from the same location) were also excluded from analysis. Each remaining specimen therefore represented a single phenological observation. Phenological models derived using herbarium-based observations...
of flowering phenology have been found to accurately predict shifts in phenological events that were observed in situ in response to climate changes (Primack et al., 2004; Miller-Rushing et al., 2006; Park and Mazer, 2018).

Species-specific models of MFD for each species were conducted using elastic net regularization, which has previously been demonstrated to be an effective method for predicting the flowering times of angiosperm taxa using herbarium specimens (Park and Mazer, 2018). For the models used by PhenoForecaster, winter and spring climate conditions (consisting of NFFD, PAS, and BFFP) at the location and DOY from which each specimen was collected were first estimated using the software package ClimateNA (Wang et al., 2016). Each species-specific phenoclimatic model was then constructed using elastic net regularization, a multivariate regression method that, rather than selecting or removing parameters in a binary fashion as with forward or backward selection, enforces parsimony by penalizing model complexity using two penalty terms: the sum of the absolute value of all parameter coefficients (L1, Eq. 1a), and the sum of all parameter coefficients squared (L2, Eq. 1b [Zou and Hastie, 2005; De Mol et al., 2009]).

\[
L_1 = \sum \| \beta \| \quad \quad (1a)
\]
\[
L_2 = \sum \| \beta^2 \| \quad \quad (1b)
\]

A penalty weighting term (\(\alpha\)) controls the degree to which model complexity is penalized. Similarly, the relative penalization of L1 versus L2 is controlled by a relative weighting term (\(\rho\)). The model for which the sum of the SSE (sum of squared errors) and the L1 and L2 penalties, modified by the two weighting terms, is minimized (C; Eq. 2) is selected as the optimal model.

\[
C = \text{SSE} + \alpha \rho \| L_1 \| + \alpha (1-\rho) \| L_2 \| \quad \quad (2)
\]

This method has substantial advantages over stepwise forward selection or backward elimination regression techniques, particularly when handling data sets in which multiple explanatory factors are likely to exhibit some degree of collinearity, such as is common in climatic data (Rawal et al., 2015). Elastic net regression has been found to generate models that remain highly stable in cases where multiple explanatory factors exhibit collinearity (Zou and Hastie, 2005), while avoiding the variance inflation that often occurs when using stepwise regression techniques (De Mol et al., 2009; Raschkla and Mirjalili, 2017).

For each angiosperm species that was represented by 100 or more specimens in our herbarium-based data set, phenological models were constructed to predict the MFD of that species from local climate conditions using the elasticCV class contained within Scikit-Learn 0.814-4 in Python, which conducts an internally cross-validated version of elastic net regularization that selects the optimal weights for the weighting terms \(\rho\) and \(\alpha\) in order to minimize both model complexity and standard error (Appendix S1).

The models used for each species in this study were constructed through iterative fitting along a regularization path,
Although many studies have examined patterns of phenological observations of mean flowering time derived from in situ pheno-
logical observations provided by the USA National Phenology Network database. The models used by PhenoForecaster predicted the timing of both in situ and herbarium-based observations of mean flowering with similar accuracy (as measured by MAE, Appendix S2). Species for which phenoclimate models produced MAE values of <15 days were considered to exhibit “good” model fits (and therefore available for selection in PhenoForecaster) by default. However, PhenoForecaster allows users to alter the MAE threshold that they consider to represent “good” model performance to accommodate cases where higher or lower predictive accuracy is required. For 186 of the 2320 species examined here, the cross-validated MAE produced by the phenological model was identical to that estimated using the collection dates of the specimens alone (i.e., without selecting climate parameters). Although these species were retained for use in PhenoForecaster, it should be noted that no climate data may be entered for these species, and the resulting predictions of flowering time consist only of a constant value reflecting an estimate of the mean observed flowering date for that species, which is not influenced by local climate conditions. Additional species will be added and models will be updated as new data or superior modeling techniques become available. Updated versions of this program will be hosted at https://labs.eemb.ucsb.edu/mazer/susan/software.

CONCLUSIONS

Although many studies have examined patterns of phenological variation in response to local climate, few tools exist for the prediction of phenological timing under novel climate conditions. PhenoForecaster provides a free, quick, and easy-to-use software package that allows researchers of any background to quickly predict the mean flowering date of angiosperm species under novel annual conditions, or at locations where the phenology of that species has not previously been observed. Its intuitive user interface and compatibility with existing spatial climate estimation packages such as ClimateNA (Wang et al., 2016) make phenological prediction easy to accomplish by researchers of any background without the need for extensive training or familiarity with phenoclimate modeling. It should be noted, however, that the accuracy of predictions by PhenoForecaster is variable and depends highly on the species selected for prediction. The expected accuracy of PhenoForecaster output, as reflected by the MAE value for that species, should be kept in mind when dealing with predicted MFD values generated by PhenoForecaster. Furthermore, these models do not account for potential heterogeneity of phenological responsiveness among populations of a given species, but instead represent mean phenological responsiveness across all available specimens for each species. These data were also based on models trained using phenological observations throughout North America only, and using derived estimates of local climate condition produced using ClimateNA; these estimates may exhibit some differences from ground-based observations of these climate parameters, or from estimates of these climate parameters derived using different methods. Thus, predictions of the phenology of these species outside of North America, or based on different sources of climate data, should be treated with caution. In addition, it should be remembered that PhenoForecaster models the timing of MFD only, and that the relationship of MFD to other phenophases, such as leaf-out, date of first flower, or date of last flower, may be highly variable among species and across climate gradients. These predictions should therefore be treated as dates on which the individuals of a given species are likely to be in flower where they have experienced a particular suite of climatic conditions, rather than as the onset or termination date of any specific phenophase. Where possible, we also recommend cross-checking predicted MFD values generated by PhenoForecaster against observed MFD values for that species, particularly when evaluating the phenology of a species under conditions that are outside of its historical range limits. Nevertheless, PhenoForecaster represents a freely available and powerful tool that allows any researcher to conduct rapid predictions of phenological timing under past, projected, or otherwise novel climate conditions.

ACKNOWLEDGMENTS

This work was supported by the National Science Foundation (DEB-1556768). Phenological data were accessed on 14 March 2017 from participants of the Consortium of California Herbaria (http://ucjeps.berkeley.edu/consortium/), SEINet (http://swbiodiversity.org/seinet/), the SERNEC Data Portal (http://ser necportal.org/portal/index.php), the Consortium of Midwest Herbaria (http://midwestherbaria.org/), the Intermountain Regional Herbarium Network (http://intermountainbiota.org), the North American Network of Small Herbaria (http://nansh.org/), the Northern Great Plains Regional Herbarium Network (http://ngpherbaria.org/), and the Consortium of Pacific Northwest Herbaria (http://pnw herbaria.org/). Individual contributing herbaria are listed in Appendix 1.
DATA ACCESSIBILITY

The PhenoForecaster source code, as well as the code used to construct the phenoclimate models, are available at https://github.com/isaacWpark/PhenoForecaster. A Windows installer package and source code, as well as a user’s manual, are available at https://labs.emb.ucsb.edu/mazer/susan/software.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

APPENDIX S1. Model parameters and predicted mean absolute error (MAE) for all species.

APPENDIX S2. Mean absolute error (MAE) of species-specific phenological model outputs in predicting (1) herbarium-based phenological observations not present in model training and (2) in situ phenological observations of mean flowering time (referred to as “in situ MAE”).

LITERATURE CITED

Aldridge, G., D. W. Inouye, J. R. K. Forrest, W. A. Barr, and A. J. Miller-Rushing. 2011. Emergence of a mid-season period of low floral resources in a montane meadow ecosystem associated with climate change. Journal of Ecology 99: 905–913.

Cannell, M. G. R., and R. I. Smith. 1986. Climatic warming, spring budburst and forest damage on trees. Journal of Applied Ecology 23: 177–191.

De Mol, C., E. De Vito, and L. Rosasco. 2009. Elastic-net regularization in learning theory. Journal of Complexity 25: 201–230.

Feldman, T. S., W. F. Morris, and W. G. Wilson. 2004. When can two plant species facilitate each other’s pollination? OIKOS 105: 197–207.

Inouye, D. W. 2008. Effects of climate change on phenology, frost damage, and floral abundance of montane wildflowers. Ecology 89: 353–362.

Miller-Rushing, A. J., R. B. Primack, D. Primack, and S. Mukunda. 2006. Photographs and herbarium specimens as tools to document phenological changes in response to global warming. American Journal of Botany 93: 1667–1674.

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

Primack, D., C. Imbres, R. B. Primack, A. J. Miller-Rushing, and P. Del Tredici. 2004. Herbarium specimens demonstrate earlier flowering times in response to warming in Boston. American Journal of Botany 91: 1260–1264.

Python Core Team. 2008. Python: A dynamic, open source programming language. Python Software Foundation Website https://www.python.org/ [accessed x month year].

Raschkla, S., and V. Mirjalili. 2017. Python machine learning, 2nd ed. Packt Publishing, Birmingham, United Kingdom.

Rathcke, B. 1988a. Flowering phenologies in a shrub community: Competition and constraints. Journal of Ecology 76: 975–994.

Rathcke, B. 1988b. Interactions for pollination among coflowering shrubs. Ecology 69: 446–457.

Rawal, D. S., S. Kasel, M. R. Keatley, and C. R. Nitschke. 2015. Herbarium records identify sensitivity of flowering phenology of eucalypts to climate: Implications for species response to climate change. Austral Ecology 40: 117–125.

Wang, T., A. Hamann, D. L. Spittlehouse, and C. Carrol. 2016. Locally down-scaled and spatially customizable climate data for historical and future periods for North America. PLoS ONE 11: e0156720.

Waser, N. M. 1978. Competition for hummingbird pollination and sequential flowering in two Colorado wildflowers. Ecology 59: 934–944.

Zou, H., and T. Hastie. 2005. Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society. Series B. Statistical Methodology 67: 301–320.

APPENDIX 1. List of contributing herbaria.

| Institution name | Herbarium name | Index Herbariorum code |
|------------------|--------------|------------------------|
| Yale University  | Peabody Museum of Natural History Herbarium | YU |
| University of Connecticut | George Safford Torrey Herbarium | CONN |
| Acadia University | E. C. Smith Herbarium | CHRB |
| University of Montreal | Marie-Victorin Herbarium | MT |
| Harvard University | Harvard University Herbaria | A, AMES, ECON, FH, GH, NEBC |
| University of New Hampshire | Albion Hodgson Herbarium | NHA |
| Drexel University | The herbarium of the Academy of Natural Sciences of Drexel University | PH |
| University of California—Berkeley | Jepson Herbarium | JEPS |
| University of California—Berkeley | Sacehen Field Station Herbarium | SCFS |
| California Polytechnic State University | California Polytechnic State University Herbarium | OBI |
| University of Santa Cruz | University of Santa Cruz Herbarium | UCSC |
| Black Hills State University | Black Hills State University Herbarium | BHSC |
| Luther College | Luther College Herbarium | LCDI |
| Minot State University | Minot State University Herbarium | MISU |
| Tarleton State University | Tarleton State University Herbarium | TAC |
| South Dakota State University | C. A. Taylor Herbarium | SDSU-SDC |
| Pittsburg State University | Theodore M. Sperry Herbarium | KSP |
| Montana State University | Billings Herbarium | MSUB |
| Sul Ross University | A. Michael Powell Herbarium | SRSC |
| Fort Hays State University | Fort Hays State University Herbarium | FHKSC |
| Utah State University | Intermountain Herbarium | USU-UTC |

(Continues)
### APPENDIX 1 (continued)

| Institution name                                           | Herbarium name                                           | Index Herbariorum code |
|------------------------------------------------------------|-----------------------------------------------------------|------------------------|
| Brigham Young University                                   | S. L. Welsh Herbarium                                     | BRY-V                  |
| Eastern Nevada Landscape Coalition                          | Eastern Nevada Landscape Coalition Herbarium              | ENLC                   |
| University of Nevada                                       | Reno Herbarium                                            | RENO-V                 |
| Natural History Museum of Utah                             | Garrett Herbarium                                         | UT-Botany              |
| Western Illinois University                                | R. M. Myers Herbarium                                     | MWI                    |
| Eastern Illinois University                                | Stover-Ebinger Herbarium                                  | EIU                    |
| Northern Illinois University                               | Northern Illinois University Herbarium                    | DEK                    |
| Morton Arboretum                                            | Morton Arboretum Herbarium                                | MOR                    |
| Chicago Botanic Garden                                     | Chicago Botanic Garden Herbarium                          | CHIC                   |
| Field Museum of Natural History                            | Field Museum of Natural History                          | F-Botany               |
| University of Wisconsin–Madison                            | University of Wisconsin–Madison, Wisconsin State Herbarium | WIS                    |
| University of Michigan                                     | University of Michigan Herbarium                          | MICH                   |
| Indiana University                                          | Deam Herbarium                                            | IND                    |
| Universidad de Sonora                                       | Universidad de Sonora Herbarium                           | USON                   |
| Centro de Investigaciones Biológicas del Noroeste, S.C.    | Observaciones Generales de Flora del Noroeste de México   | RHNM                   |
| Instituto Politécnico Nacional–Unidad Durango              | Herbario del Instituto Politécnico Nacional–Unidad Durango | CIDIR                  |
| University of California–Riverside                         | University of California–Riverside Herbarium              | UCR                    |
| San Diego State University                                 | San Diego State University Herbarium                      | SDSU                   |
| Granite Mountains Desert Research Center                    | Granite Mountains Desert Research Center                  | GMDRC                  |
| University of South Carolina                                | A. C. Moore Herbarium                                     | USCH                   |
| Auburn University                                           | John D. Freeman Herbarium                                 | AUJ                    |
| Clemson University                                          | Clemson University Herbarium                              | CLEMS                  |
| Eastern Kentucky University                                 | Ronald L. Jones Herbarium                                 | EKY                    |
| College of William and Mary                                 | College of William and Mary Herbarium                     | WILI                   |
| Appalachian State University                               | I. W. Carpenter, Jr. Herbarium                            | BOON                   |
| University of North Carolina                                | University of North Carolina Chapel Hill Herbarium        | NCU                    |
| University of Memphis                                       | University of Memphis Herbarium                           | MEM                    |
| Mississippi State University                                | Mississippi State University Herbarium                    | MISSA                  |
| University of Mississippi                                   | Thomas M. Pullen Herbarium                                | MISS                   |
| University of Southern Mississippi                         | University of Southern Mississippi Herbarium              | USMS                   |
| Mississippi Museum of Natural Science                       | Mississippi Museum of Natural Science Herbarium            | MMNS                   |
| Marshall University                                         | Marshall University Herbarium                              | MUHW                   |
| Longwood University                                         | Harvill-Stevens Herbarium                                 | FARM                   |
| Western Carolina University                                 | Western Carolina University Herbarium                     | WCUH                   |
| Northern Kentucky University                                | John W. Thieret Herbarium                                 | KNK                    |
| Salem College                                               | Salem College Herbarium                                    | SC                     |
| Troy University                                             | Troy University Herbarium                                 | TROY                   |
| Arizona State University                                    | Arizona State University Herbarium                        | ASU-Plants             |
| University of Arizona                                       | University of Arizona Herbarium                           | ARIZ                   |
| Desert Botanical Garden                                     | Desert Botanical Garden Herbarium                         | DES                    |
| Northern Arizona University                                 | Deaver Herbarium                                           | ASC                    |
| Navajo Nation Department of Fish and Wildlife               | Navajo Nation Herbarium                                   | NAVA                   |
| Grand Canyon National Park                                  | Grand Canyon National Park Herbarium                      | GCNP                   |
| University of New Mexico                                    | University of New Mexico Herbarium                        | UNM-Vascular Plants    |
| Western New Mexico University                               | Dale A. Zimmerman Herbarium                               | SNM                    |
| Museum of Northern Arizona                                  | Museum of Northern Arizona Herbarium                      | MNA                    |
| Gila National Forest                                        | Gila National Forest Herbarium                            | USFS-GILA              |
| Arizona Western College                                     | Arizona Western College Herbarium                          | AWC                    |
| Natural History Institute                                   | Natural History Institute Herbarium                        | NHI                    |