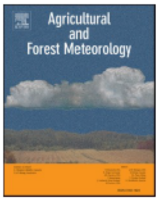
Contents lists available at ScienceDirect

Agricultural and Forest Meteorology

journal homepage: www.elsevier.com/locate/agrformet



Predicting spring phenology in deciduous broadleaf forests: NEON phenology forecasting community challenge

Kathryn I. Wheeler[a,b,c,d,*], Michael C. Dietze[a], David LeBauer[e], Jody A. Peters[f], Andrew D. Richardson[g,h], Arun A. Ross[i], R. Quinn Thomas[j,k], Kai Zhu[l,m], Uttam Bhat[n], Stephan Munch[o], Raphaela Floreani Buzbee[p], Min Chen[q], Benjamin Goldstein[p], Jessica Guo[e], Dalei Hao[r], Chris Jones[s], Mira Kelly-Fair[a], Haoran Liu[q], Charlotte Malmborg[a], Naresh Neupane[t], Debasmita Pal[i], Vaughn Shirey[t], Yiluan Song[u], McKalee Steen[p], Eric A. Vance[v], Whitney M. Woelmer[k], Jacob H. Wynne[k,w], Luke Zachmann[x]

[a] Department of Earth and Environment, Boston University, 685 Commonwealth Avenue, Boston, MA 02215, United States
[b] Cooperative Programs for the Advancement of Earth System Science, University Corporation for Atmospheric Research, Foothills Laboratory Bldg 4, 3300 Mitchell Lane, Boulder, CO 80301, United States
[c] National Oceanic and Atmospheric Administration, Foothills Laboratory Bldg 4, 3300 Mitchell Lane, Boulder, CO 80301, United States
[d] Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, 15 Vassar St, Cambridge, MA 02142, United States
[e] Arizona Experiment Station, University of Arizona, 1230N Cherry Ave, Tucson, AZ 85721, United States
[f] Department of Biological Sciences, University of Notre Dame, 100 Galvin, Notre Dame IN 46556, United States
[g] School of Informatics, Computing and Cyber Systems, Northern Arizona University, 1295 Knoles Dr. Flagstaff, AZ 86011, United States
[h] Center for Ecosystem Science and Society, Northern Arizona University, PO Box 5620, Flagstaff, AZ 86011, United States
[i] Department of Computer Science and Engineering, Michigan State University, 428 South Shaw Lane, East Lansing, MI 48824, United States
[j] Department of Forest Resources and Environmental Conservation, Virginia Polytechnic Institute, 310 West Campus Drive, Blacksburg, VA 24061, United States
[k] Department of Biological Sciences, Virginia Tech, 926 West Campus Drive, Blacksburg, VA 24061, United States
[l] Institute for Global Change Biology, University of Michigan, 440 Church Street, Ann Arbor, MI 48109, United States
[m] School for Environment and Sustainability, University of Michigan, 440 Church Street, Ann Arbor, MI 48109, United States
[n] Institute of Marine Sciences, University of California Santa Cruz, 1156 High Street, Santa Cruz, CA 95060, United States
[o] Department of Applied Mathematics, University of California, Santa Cruz, 1156 High Street, Santa Cruz, CA 95060, United States
[p] Department of Environmental Science, Policy, and Management, University of California, Berkeley, 4004 Valley Life Sciences Building University of California, Berkeley, CA 94720, United States
[q] Department of Forest and Wildlife Ecology, University of Wisconsin-Madison, 1630 Linden Dr., Madison, WI 53706, United States
[r] Atmospheric, Climate, and Earth Sciences Division, Pacific Northwest National Laboratory, 902 Battelle Blvd, Richland, WA 99354, United States
[s] Center for Geospatial Analytics, North Carolina State University, 5112 Jordan Hall, 2800 Faucette Drive, Raleigh, NC 27695, United States
[t] Department of Biology, Georgetown University, 37th and O streets, NW, Washington, DC 20057, United States
[u] Department of Environmental Studies, University of California, 1156 High St, Santa Cruz, CA 95064, United States
[v] Department of Applied Mathematics, University of Colorado Boulder, 1111 Engineering Drive, Boulder, CO 80309, United States
[w] Department of Microbiology, Oregon State University, 2820 SW Campus Way, Corvallis, OR 97331, United States
[x] Conservation Science Partners, Inc., 11050 Pioneer Trail, Suite 202, Truckee, CA 96161, United States

ARTICLE INFO

Keywords:
Phenology
Ecological forecasting
Deciduous broadleaf
Budburst

ABSTRACT

Accurate models are important to predict how global climate change will continue to alter plant phenology and near-term ecological forecasts can be used to iteratively improve models and evaluate predictions that are made a priori. The Ecological Forecasting Initiative's National Ecological Observatory Network (NEON) Forecasting Challenge, is an open challenge to the community to forecast daily greenness values, measured through digital images collected by the PhenoCam Network at NEON sites before the data are collected. For the first round of the

Abbreviations: CRPS, Continuous ranked probability Score; DB, Deciduous broadleaf; DN, Digital number; DOY, Day of year; EFI, Ecological forecasting initiative; GAM, Generalized additive model; G_{CC}, Green chromatic coordinate; GDD, Growing degree day; GEFS, Global ensemble forecast system; MAT, Mean annual temperature; ML, Machine learning; MODIS, Moderate resolution imaging spectroradiometer; NEON, National ecological observatory network; ROI, Region of interest.

[*] Corresponding author at: Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, 15 Vassar St, Cambridge, MA 02142, United States.

E-mail address: kwheelerecology@gmail.com (K.I. Wheeler).

https://doi.org/10.1016/j.agrformet.2023.109810
Received 11 August 2023; Received in revised form 5 October 2023; Accepted 8 November 2023

Community challenge
Forests

challenge, which is presented here, we forecasted canopy greenness throughout the spring at eight deciduous broadleaf sites to investigate when, where, and for what model type phenology forecast skill is highest. A total of 192,536 predictions were submitted, representing eighteen models, including a persistence and a day of year mean null models. We found that overall forecast skill was highest when forecasting earlier in the greenup curve compared to the end, for shorter lead times, for sites that greened up earlier, and when submitting forecasts during times other than near budburst. The models based on day of year historical mean had the highest predictive skill across the challenge period. In this first round of the challenge, by synthesizing across forecasts, we started to elucidate what factors affect the predictive skill of near-term phenology forecasts.

1. Introduction

Plant phenology is a primary ecological indicator of climate change (Parmesan and Yohe, 2003) and impacts a variety of ecosystem processes including surface roughness, albedo, canopy conductance, and carbon dioxide and water fluxes (Richardson et al., 2013). Realistic representations of plant phenology are crucial for reliable global carbon and water cycle predictions in climate models (Stockli et al., 2008). In particular, the timing of spring phenology is advancing earlier (Piao et al., 2019) and influences other phenological transitions, such as senescence (Keenan and Richardson, 2015). Given this, we want to be able to anticipate future changes in phenology by assessing how well models perform in the near-term.

One important plant functional type that is changing is cold-deciduous plants (Piao et al., 2019), which enter dormancy through shedding leaves in cold conditions. In the spring, they experience budburst to break dormancy and then increases in total leaf area through expansion and unfolding (Chuine and Regniere, 2017). Budburst requires both chilling during the first stage of dormancy (endodormancy) and warming during the second stage (ecodormancy; Chuine and Regniere, 2017). Since monitoring the switch from endodormancy to ecodormancy is challenging, we might expect that predicting canopy greenness around budburst is harder than other parts of the spring, but it is unclear if this is true. Even though we understand spring physiological mechanisms, there exists a large variation in the types of phenology models (Chuine and Regniere, 2017; Piao et al., 2019). Additionally, since the drivers differ between warm and cold regions (Moon et al., 2021; Zohner et al., 2016) and warmer regions tend to budburst earlier than colder regions, site predictability likely differs based on greenup timing, but it is unclear how. One way to improve our ability to model phenology is through ecological forecasting.

Near-term ecological forecasting has societal and scientific benefits. By creating an iterative feedback loop on learning and model improvement, it accelerates our scientific understanding, and by withholding yet-to-be-collected future data for validation (Dietze, 2017; Dietze et al., 2018), it makes models more robust by reducing the possibility of overfitting. Therefore, making and evaluating forecasts can help reveal phenology predictability and elucidate which types of models have the highest skill in the near-term. Additionally, unlike studies that predict changes to phenology in 2100 under climate change scenarios (*e.g.,* Archetti et al., 2013; Delpierre et al., 2009; Keenan and Richardson, 2015; Lebourgeois et al., 2010; Xie et al., 2018), near-term forecasting allows for the rapid and iterative testing of models and hypotheses against new observations. Improving near-term phenology forecasts has benefits ranging from informing scientists of data collection times, optimizing land management activities, and improving weather forecasts (Morisette et al., 2009; Xue et al., 1996). Furthermore, since canopy greenness data can have a low latency (*e.g.,* less than a day), phenology forecasts are not subject to data reporting delays, which are common in other forecasted ecological systems (Johansson et al., 2019). Thus, phenology is an ideal system for executing ecological forecasting and testing forecasting theory.

Previous efforts to forecast phenology have spanned spatial and temporal scales from the greenup of individual species made based on *in situ* observations (Gerst et al., 2021) to predictions of land surface phenology using vegetation indices derived from satellite imagery (*e.g.,* Neupane et al., 2022; Xu et al., 2021). Additional examples include the large-scale forecasting efforts by the United States National Phenology Network (Crimmins, 2020) and the automated species-level forecast system of Taylor and White (2020). Existing forecasts typically focus on forecasting the timing of specific phenological transition dates instead of daily phenological conditions. However, the underlying seasonal processes of phenological development are typically continuous and dynamic, meaning that the phenological condition tomorrow is based on that of today.

In addition to improving the representation of the physiological process, representing phenology as continuous also allows for iterative data assimilation approaches that update predictions continuously with new phenological observations (Viskari et al., 2015). By providing the potential to assimilate more data that is closer temporally to when is being forecasted, this likely cause forecasts to become more accurate as lead time (*i.e.,* difference between the date forecasted and the date the forecast was submitted) decreases. Additionally, forecast skill should be higher for shorter lead times because of the influence of lead time on meteorological forecasts accuracy. Thus, there is an unmet opportunity to improve our understanding of phenological processes and how dynamic models compare to other model classifications by forecasting spring green-up as a continuous process.

Unlike previous phenology forecasting efforts that occurred without a common community framework, an open community challenge with a clear set of guidelines and shared cyberinfrastructure and data allows for comparisons across forecast models and provides insight into the predictability of phenology outside of one specific model or team of people. Through providing a common pipeline to increase the ease of running forecasts, it also encourages more people to try forecasting, bringing with them different and creative perspectives. This community framework should help speed up the process of model development and forecasting phenology in the near-term.

In response to this need, the Ecological Forecasting Initiative s (EFI) Research Coordination Network is hosting the NEON (National Ecological Observatory Network) Ecological Forecasting Challenge, an open community forecasting challenge where a design team provides infrastructure and guidance and teams in the ecological and related communities can submit forecasts of NEON data (Thomas et al., 2023a). In Round 1 of the Phenology Forecast Challenge in 2021, teams were asked to forecast daily canopy greenness at eight cold-deciduous broadleaf sites within NEON. We hypothesized that (H1) forecasts would improve as lead time decreases; (H2) the start of greenup (*i.e.,* budburst) would be the hardest part of the curve to forecast; (H3) similarly, forecasts submitted right before budburst would have the lowest predictive power as they are forecasting the greenup curve; (H4) dynamic models that assimilate new phenology data would perform better than those that do not; and (H5) differences in predictability between sites would be explained by differences in the timing of greenup. The answers to these hypotheses have important ramifications for understanding what impacts phenology forecast skill and the maturity of our community s modeling efforts.

2. Methods

2.1. NEON phenology forecasting challenge description

The NEON Phenology Forecasting Challenge is an open challenge that teams can submit to and join at any time, using multiple models if desired (Thomas et al., 2023a). For Round 1, teams were tasked with forecasting daily PhenoCam G_{CC} at eight NEON sites throughout spring for 35 days into the future for each submission day. All models had at least one team member who participated as a co-author. The design team provided a target file of previous PhenoCam data for each site that was updated daily with new PhenoCam data. Forecasts were to be submitted by 6 pm ET each day with the first day of the forecast being the submission day (*e.g.,* a submission on 1 February 2021 included forecasts for 1 February 2021–7 March 2021). A small number of teams from university courses were permitted to submit late forecasts provided no data beyond the forecast start date was used. Forecasts longer than 35 days were filtered out in this initial analysis. Submissions had to include uncertainty estimates and be submitted in the Ecological Forecasting Initiative forecast standard (Dietze et al., 2023).

2.2. Site selection and description

We selected eight temperate sites within NEON (Table 1 and Fig. 1) that represented seven different ecoclimatic domains and that included deciduous broadleaf plants within view of a PhenoCam. Sites had two to four years of PhenoCam data prior to the start of the Challenge, but many sites had longer-term PhenoCams located nearby (*e.g.,* Harvard Forest) that could be used in calibration. Additionally, some sites (*e.g.,* Bartlett) were selected for overlap with other forecasting challenges (terrestrial fluxes, aquatic temperature and dissolved oxygen, tick populations, and beetle fluxes) within the EFI NEON Ecological Forecasting Challenge in Round 1 (Thomas et al., 2023a).

2.3. Phenology data: PhenoCam

To monitor canopy greenness, we used NEON PhenoCams, which are digital cameras that take regular repeated images of plant canopies as part of the PhenoCam Network (Seyednasrollah et al., 2019). The low latency of PhenoCam data (less than one day), provides an opportunity to evaluate forecasts in real-time. NEON's PhenoCams were installed following the standard PhenoCam Network deployment protocol (Richardson et al., 2018). Each camera (NetCam SC IR, StarDot Technologies, Buena Park, CA, USA) was configured using automated scripts (the PhenoCam Installation Tool) to ensure consistency in settings such as exposure and white (color) balance, as well as image and metadata acquisition and transmission. Multiple cameras are deployed at each NEON site; for this Challenge, the data were derived from the top-of-tower cameras.

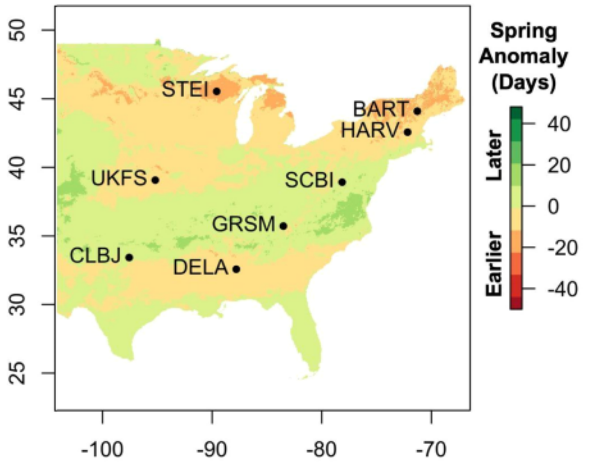


Fig. 1. Locations of selected sites and the National Phenology Network's Historical Annual Spring Indices Anomaly for First Leaf product during the study year of 2021 compared to the 1991–2020 average (USA National Phenology Network, 2017). For full site descriptions and names see Table 1. Bartlett (BART), Harvard Forest (HARV), Steigerwaldt (STEI), Dead Lake (DELA), and University of Kansas (UKFS) likely experienced earlier than site-average greenup and Lyndon B. Johnson (CLBJ), Great Smokies (GRSM), and Smithsonian (SCBI) likely experienced later than average greenup.

Each NEON camera is set to record an image every 15 min. Quantitative image analysis consists of several steps. First, an appropriate "region of interest" (ROI) is defined for the camera, corresponding to the area within each digital image for which color information will be extracted. Second, images are sequentially read in, and the frequency distribution of the pixel values (pixel value is an 8-bit digital number, or DN) for each color channel (red, green, and blue) is characterized for the ROI in each image. Third, a normalized vegetation index, the green chromatic coordinate (G_{CC}), is calculated:

$$G_{CC} = G_{DN} / (R_{DN} + G_{DN} + B_{DN}), \tag{1}$$

G_{CC} has been shown to be highly effective at suppressing variation due to external factors, such as scene illumination (weather and atmospheric effect), and maximizing the underlying phenological signal. G_{CC} is

Table 1
Summary of site characteristics.

Site Name	Site (and PhenoCam) ID	Latitude	Longitude	MAT (°C)	Number of Years	Reported Dominant DB Species
Harvard Forest, MA (HARV)	NEON.D01.HARV.DP1.00033	42.537	−72.173	7.15	4	*Quercus rubra*
Bartlett Experimental Forest, NH (BART)	NEON.D01.BART.DP1.00033	44.0639	−71.287	6.1	4	*Fagus grandifolia, Acer rubrum, Betula papyrifera*
Smithsonian Conservation Biology Institute, VA (SCBI)	NEON.D02.SCBI.DP1.00033	38.893	−78.140	11.8	4	*Liriodendron tulipifera, Juglans nigra*
Steigerwaldt Land Services, WI (STEI)	NEON.D05.STEI.DP1.00033	45.509	−89.586	4.95	3	*Populus tremuloides, Acer rubrum*
The University of Kansas Field Station, KS (UKFS)	NEON.D06,UKFS.DP1.00033	39.040	−95.192	12.65	2	*Symphoricarpos orbiculatus, Celtis occidentalis, Carya ovata*
Great Smoky Mountains National Park, TN (GRSM)	NEON.D07.GRSM.DP1.00033	35.689	−83.502	12.65	3	*Liriodendron tulipifera, Acer rubrum, Acer pensylvanicum*
Dead Lake, AL (DELA)	NEON.D08,DELA.DP1.00033	32.542	−87.804	17.9	4	*Celtis laevigata, Ligustrum sinense, Liquidambar styraciflua*
Lyndon B. Johnson National Grassland, TX (CLBJ)	NEON.D11.CLBJ.DP1.00033	33.401	−97.570	17.65	3	*Quercus marilandica, Schizachyrium scoparium*

Note: MAT refers to Mean Annual Temperature and comes from the Daymet (Thornton et al., 2017) estimation provided by PhenoCam at https://phenocam.nau.edu/. DB refers to deciduous broadleaf. The number of years indicates how many years before 2021 each camera started regularly collecting data.

calculated for each image, but images obtained when the solar elevation was less than 5 , or images that are too bright or too dark (Klosterman et al., 2014; Toomey et al., 2015) are excluded. Finally, a daily value of G_{CC} is calculated using the 90th quantile approach described by Sonnentag et al. (2012) and, for the challenge, the standard deviation of the 90th quantile value of G_{CC} was estimated through bootstrapping. Data are processed and posted daily. More information about data processing is available in Seyednasrollah et al. (2019).

2.4. Forecasted meteorological data: global ensemble forecast system

While use was not required, the design team provided teams with site-specific meteorological forecasts extracted from National Oceanic and Atmospheric Administration s Global Ensemble Forecast System (GEFS; Li et al., 2019; https://www.nco.ncep.noaa.gov/pmb/products/gens/). The midnight UTC forecast was selected because it contained 30 ensemble members that extended 35 days into the future; each ensemble member was temporally downscaled to one hour temporal resolution. GEFS variables include air temperature, air pressure, wind speed, precipitation, downwelling longwave (thermal) radiation, downwelling shortwave (solar) radiation, and relative humidity. Teams could access an S3 bucket with the GEFS forecasts online via the Amazon Web Services Application Programming Interface, or via the 'neon4cast R package (Boettiger and Thomas, 2022) as part of the Challenge cyberinfrastructure.

2.5. Null models

Submitted forecasts were compared to two null models. The first is the persistence, or random walk, model, which assimilates new PhenoCam data daily and predicts the next day s G_{CC} value as the current day s plus normally distributed error. The second is the day of year (DOY) historical mean of all previous years that were available for each PhenoCam, which consists of the mean and standard deviation averaged over that PhenoCam s previous years G_{CC} values for each DOY.

2.6. Modeling teams

Thirteen distinct teams submitted forecasts from eighteen models, including the two null models (Table 2). More detailed model descriptions are given in the Supplementary Materials. To assess H4, we classified models into distinct types, focusing specifically on two high-level factors: (1) whether the approach made use of time-varying covariates (*e.g.,* weather forecast) and (2) whether the model was dynamic (prediction of G_{CC} tomorrow is a function of G_{CC} today). For the analyses, the eighteen models were thus grouped into five different types: DOY Mean, persistence, static (does not use covariates or the previous G_{CC} state), covariate (uses covariates but not the previous state), and dynamic (includes previous state). While the persistence model is technically a dynamic model, we excluded it from the dynamic class for this analysis to evaluate it separately as a null model. Further assessing differences in model types was limited due to the large variety of modeling approaches employed.

Table 2
Summary of models.

Team ID	Approach	Model Type: DOY Mean, Persistence, Covariate, Dynamic, or Static	Covariates (not including previous G_{CC} values)	Uncertainties included (Driver, Initial condition, parameter, process, and observational)	References
CSP_Gwave	Statistical	Covariate	Site level summaries of precipitation, temperature, and their interaction, latitude, and long-run greenness (from MODIS)	Parameter, process, and observational	None
CU_Pheno	Process	Dynamic	GDD, maximum G_{CC}	Driver and initial condition	None
DALEC_SIP	Process	Dynamic	GDD	None	(Bloom and Williams, 2015; Chen et al., 2016; Wu et al., 2021; Yang et al., 2016; Zeng et al., 2018)
EFI_U_P	Process	Covariate	DOY	Parameter, and observational	None
Fourier	Statistical	Static	DOY	Observational	None
greenbears_gams	Statistical	Static	DOY	Parameter	(Wood, 2017)
greenbears_par	Statistical	Covariate	DOY, historical photosynthetically active radiation	Parameter	(Wood, 2017)
greenbears_stl	Statistical	Static	DOY	Parameter	(Wood, 2017)
PEG	Statistical	Dynamic	DOY	Parameter	(Hyndman and Khandakar, 2008)
PEG_RFR	ML	Dynamic	DOY	Observational	(Breiman, 2001; Pedregosa et al., 2011)
PEG_RFR0	ML	Dynamic	DOY	Observational	(Breiman, 2001; Pedregosa et al., 2011)
PEG_RFR2	ML	Covariate	Maximum and minimum temperature, radiation, and precipitation	Driver and observational	(Breiman, 2001; Pedregosa et al., 2011)
PhenoPhriends	Process	Dynamic	Temperature	Driver, initial condition, parameter, and process	None
Team_MODIS	Statistical	Covariate	Growing degree days, MODIS greenness onset	Driver, initial condition, observational	(Neupane et al., 2022)
GPEDM	Statistical	Dynamic	Daily mean temperature, daily total precipitation	Driver, initial condition, parameter, process, and observational	(Munch et al., 2017)
VT_Ph_GDD	Process	Covariate	GDD	Driver, parameter, process, and observational	None
DOY Mean	Statistical	DOY Mean	None	Initial condition	
Persistence	Statistical	Persistence	None	Initial condition and process	

Note: DOY refers to day of year, ML refers to machine learning, G_{CC} refers to the green chromatic coordinate, GDD refers to Growing Degree Day, and MODIS refers to Moderate Resolution Imaging Spectroradiometer. Driver is uncertainty in model drivers, covariates, and exogenous scenarios; initial condition refers to the uncertainty in the initialization of state variables (G_{CC} at time 0); parameter is uncertainty in model parameters and coefficients; process is the dynamic uncertainty in the process model attributable to both model misspecification and stochasticity; and observational is the uncertainty in the observations of the output variables (G_{CC}).

2.7. Statistical analyses

Analyses were limited to the period of February June 2021 to capture the entire transition from dormant to full canopy states across all eight sites. We assessed each forecast s skill based on the Continuous Ranked Probability Score (CRPS), which was calculated using the *crps_sample* function in the ‘scoringRules R package (Jordan et al., 2019). CRPS is a model assessment metric that scores based on both accuracy (mean absolute error) and precision (ensemble spread), and thus, has the same units as the variable being scored (in this case, G_{CC}, which is unitless as a ratio). Forecasts CRPS values and how they compared to the null models were available in real-time on the Challenge s public dashboard. Forecasts were accepted every day during this period, though not all teams submitted forecasts each day.

Since two of our hypotheses (H2 and H3) involved the skill of forecasts relative to when greenup occurred at each site, we calculated the start, middle, and end of spring (defined as 15 %, 50 %, and 85 % greenup, respectively) for each PhenoCam site using the function *ElmoreFit* in the R package ‘phenopix (Elmore et al., 2012; Filippa et al., 2020). Additionally at each site and for each model, we calculated how many total days before each transition date the forecasted G_{CC} values had a lower (*i.e.,* better) CRPS than the DOY Mean null. To investigate the forecasted year greenup anomalies, we computed the average timing of transition dates (15 %, 50 %, and 85 %) for each site. It is important to note, though, that this average was done over a small sample size (two to four years depending on the site) and does not necessarily represent a robust estimate of historical greenup. Therefore, we also assessed the forecasted year deviations using the National Phenology Network s Historical Annual Spring Indices Anomaly for First Leaf product for 2021 (USA National Phenology Network, 2017).

Statistical analyses focused on understanding the impacts of factors on forecast predictability: site, model, model type, lead time, and phenodate (either of submission date or forecasted date). Lead time was defined as the difference between the date forecasted and the date the forecast was submitted. Phenodate was defined here as days relative to the date of 15 % greenup for each site, with the sign convention of negative phenodates being days before this threshold.

Since we expect the relationship between CRPS and either lead time or phenodate to be nonlinear we analyzed the full set of predictions using Generalized Additive Models (GAMs) using the R ‘mcgv package (Wood, 2022; R Core Team, 2022; Version 4.2.2). Specifically, lead time and phenodate were modeled using thin plate regression splines using the default number of knots (n 10). In addition to providing statistical tests and high-level summaries of each factor, GAMs also help to account for differences in model submission dates across teams by correcting CRPS for phenodate and lead time.

Analyses started with an overall model that included linear terms for site and team and additive spline terms for lead time and either phenodate$_{\text{forecasted}}$ (H2) or phenodate$_{\text{submitted}}$ (H3). To assess H4, we used the model class effects (reference class DOY Mean) in the overall model created with phenodate$_{\text{submitted}}$. To assess if there was a significant relationship between when models started submitting forecasts and their overall skill, we performed a linear regression of the model effect in the overall model created from phenodate$_{\text{submitted}}$ *versus* the day of the Challenge the model first submitted a forecast. Additionally for models that had higher predictive skill (lower CRPS) than the DOY Mean, we assessed if the CRPS values were lower because of lower uncertainties in the forecasts through a Welch Two-Sample t-Test. Similarly to assess differences in predictability between sites (H5), we used the site effects (reference site Bartlett (BART)) in the overall GAM. Bartlett was used as the reference site because the ‘mcgv R package uses the first alphabetically as a default. To assess H5 and how overall seasonality affected skill, these site effects were also regressed against the dates of 15 %, 50 %, and 85 % greenup and the forecasted year timing anomalies. Furthermore, to assess whether models performed better at some sites than others we also refit the GAM with a site-by-model interaction term, which was visualized as a barplot.

In addition to the overall GAMs, to answer H1 H3 we also assessed the impact of model and model type on lead time and phenodate responses by fitting a series of independent GAMs for each model and model type. Two models forecasts, EFI_U_P and greenbears_stl, were excluded from the model-specific analyses because there were not enough submissions to fit the GAMs independently (Fig. S1). Attempts to include separate spline responses by model or model type within the overall GAM failed to converge so we do not provide an overall statistical test on these interaction terms, but instead, focus on visualizing responses.

We visualized the responses through GAM response surfaces of predicted CRPS over lead times (0 35 days) and phenodates (80 days before 15 % greenup 40 days after) for different sites, models, and model classes. When we varied phenodates, the GAM response surface represents predicted CRPS across all lead times. Similarly when we varied lead times, the response represents the predicted CRPS across all phenodates. Using response surfaces allowed us to predict CRPS for each site, model, and model class across different lead times and phenodates even if each combination did not occur in the actual forecasts (*i.e.,* not all models forecasted all dates, but we could use the fitted GAMs to predict what CRPS would have been). Additionally to further assess differences in site predictability, we regressed the maximum of the GAM response surface of varied phenodates$_{\text{forecasted}}$ for each site against greenup length.

3. Results

3.1. Forecast submissions

Overall, 192,536 individual predictions (forecast of G_{CC} by one model on one submission date for one site and one forecasted day) were submitted for eighteen models from thirteen teams, including the two null models provided (DOY Mean and persistence). All models submitted a forecast for at least one day that fell within the spring greenup period (*i.e.,* between 15 % and 85 % greenup) for at least one site, but the date of first submission and frequency varied greatly (Fig. 2), ranging from submitting on all days of the challenge to only submitting on one day (Supplementary Fig. S1). Additionally as the challenge period progressed, the average number of submissions each day increased until around mid-May 2021 (Fig. S2). Classifications of models varied and we had three, six, and seven models that were static, covariate, and dynamic, respectively, in addition to the null DOY Mean and persistence models (Table 2).

3.2. Example set of forecasts

To give an example of forecasts during the greenup period and how they differed between models, we provide an example of submitted forecasts for one site, Harvard Forest, submitted on one reference datetime, 11 May 2021, for the 35-day horizon the challenge requested (Fig. 3). We highlight Harvard Forest because it is well-known in the ecology community and finished greening up last, allowing more teams to forecast it. We chose 11 May 2021 as an example because it was right before greenup started and had the largest number of forecasts submitted. All models that submitted forecasts on this day, other than the persistence null model, predicted greenup would occur during the next 35 days (Fig. 3). Forecasted greenup timing, rate, and uncertainty all varied between teams (Fig. 3). For example, the start of greenbear_par s forecasted greenup curve was close to the observed start, but their forecasted greenup occurred slower than the actual greenup. Additionally, DALEC_SIP forecasted too early of a start and too low post-greenup G_{CC}. Furthermore, PEG_RFR forecasted post-greenup G_{CC} correctly, but predicted greenup later than it occurred. Additional examples at other sites are available in Fig. S3.

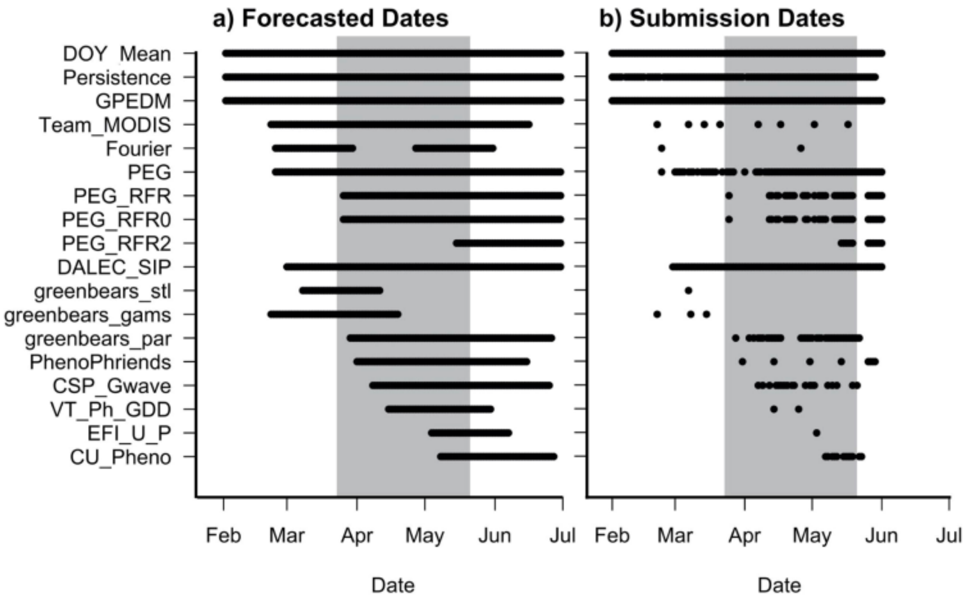


Fig. 2. The specific days that each model forecasted (a) and the days that each team submitted forecasts on (b). The period where at least one site was between 15 % and 85 % greenup is indicated with shading.

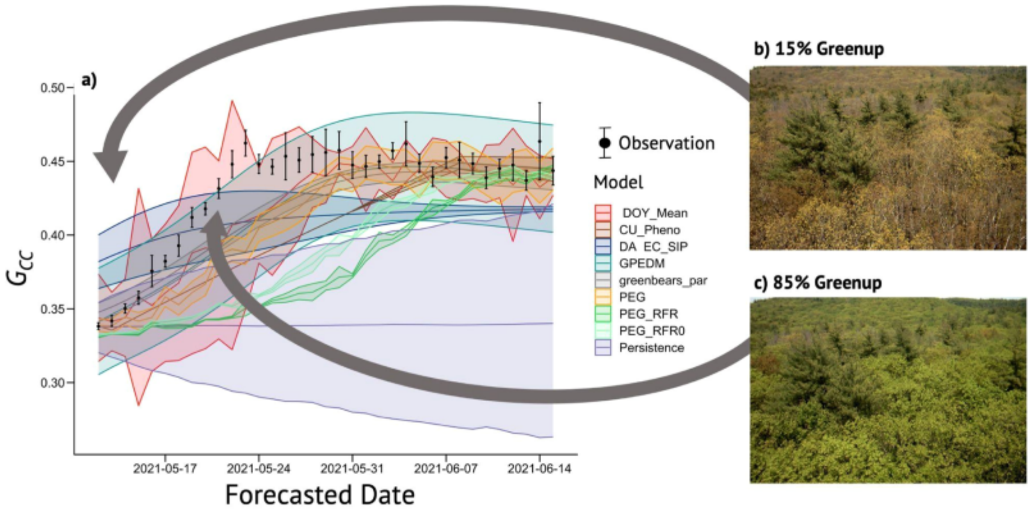


Fig. 3. a) An example of forecasted greenness values (G_{CC}) submitted by teams on 11 May 2021 for Harvard Forest. Observed G_{CC} values are given in black points with standard deviations indicated with bars. Of the teams that submitted on this date (including DOY Mean), most predicted a greenup curve during this time period. b) PhenoCam image on 15 May 2021 (date of 15 % greenup; Milliman et al., 2019). c) PhenoCam image on 21 May 2021 (date of 85 % greenup; Milliman et al., 2019).

3.3. *H1: changes in forecasts with lead time*

When considering the effect of lead time on forecast skill, the persistence, dynamic, overall, and static model types all show the expected pattern of error increasing with lead time (H1), but form a gradient, from fastest to slowest, in the rate at which error increased (Fig. 4). The DOY Mean forecast error does not vary with lead time, which is expected as this forecast is based solely on previous years data and does not change with lead time. Finally, the covariate model class exhibits a decrease in error as lead time increases.

In the GAMs fit to individual models CRPS values (Fig. 4c), most models also followed the expected pattern of increasing error with lead time (H1), while Team_MODIS, greenbears_par, and PEG_RFR2 show a similar pattern to the covariate group of a slight decrease in error with lead time. Fourier, PEG_RFR, and PEG_RFR0 all showed a pattern of error increasing to a maximum around day 20–25 before declining slightly, while in VT_Ph_GDD error declined slightly as lead time increased over the first week (similar to Team_MODIS, greenbears_par, and the covariate group), before switching to the expected pattern of an increase in error with lead time. On average, error increased fastest with lead time for the persistence (random walk) null model, which was also the worst overall performing model, suggesting that all models were consistently more skillful than a persistence null. That said, CU_Pheno and VT_Ph_GDD both had specific periods where the rate at which their error increased was more rapid than the persistence null. Specifically, CU_Pheno exhibited a rapid increase in error over the first five days before asymptoting over the remainder of the 35-day forecast, while error in VT_Ph_GDD increased more rapidly than persistence over days 13–24. Examples of how forecasted G_{CC} and skill of transition dates change with lead time are shown in Fig. S4 and S5.

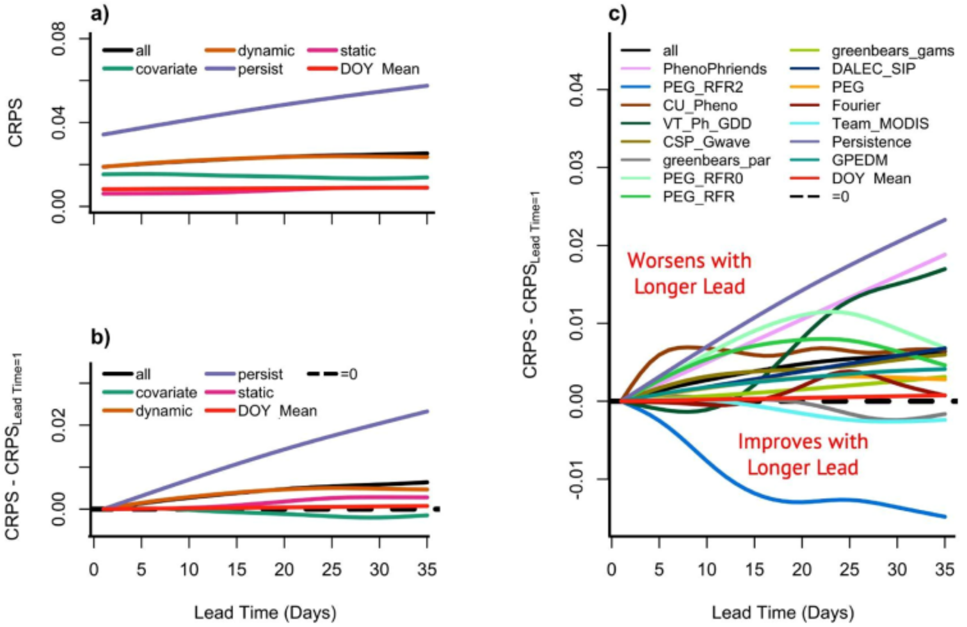


Fig. 4. a) Generalized Additive Model response surfaces of Continuous Ranked Probability Score (CRPS) as a function of lead time (*i.e.*, difference between the date forecasted and the date the forecast was submitted) across models and forecast start dates (black line) or separated by model type. b) Same as top but focusing on the change in CRPS relative to the shortest lead time (time=0)), which emphasizes changes in predictability with lead time rather than absolute skill. c) Change in CRPS with lead time for individual models.

3.4. H2: changes in forecasts with phenodate forecasted

Contradicting H2, forecast skill on average across all models, lead times, and sites was highest (CRPS lowest) when forecasting G_{CC} on days prior to greenup and lowest around 85 % greenup. Specifically, forecast skill was at a minimum on average 14 days after 15 % of greenup (black line in Fig. 5) and 0.75 days before 85 % greenup. The number of days after 15 % greenup that predictability was the worst varied between sites with Steigerwaldt, Harvard, and Bartlett reaching worst predictability first and Great Smokies and Lyndon B. Johnson last (Fig. 5). The peak magnitude of CRPS GAM response surfaces over varying phenodate$_{\text{forecasted}}$ for each site correlated with how quick greenup occurred ($R^2 = 0.87$, *p*-value = 0.0007; *F*-statistic = 40.16; degrees of freedom = 7), with the sites that greened up fastest having worse predictability.

3.5. H3: predictability based on phenodate submitted

In addition to the predictability being lowest when the *forecasted day* (phenodate$_{\text{forecasted}}$) was around 85 % greenup (Section 3.4) and supporting H3, the overall GAM showed that the predictive power was lowest for days when the forecasts were *submitted* (phenodate$_{\text{submitted}}$) right before budburst (15 % greenup). The GAM response surface starts from a constant low CRPS during the dormant season, begins to rise starting about a month before greenup, peaks four days prior to 15 % greenup (*i.e.*, phenodate$_{\text{submitted}}$ = 0), and declines to a new, higher, summer asymptote approximately three weeks after greenup (Fig. 6a). Across all model classes, the pattern in CRPS *versus* phenodate$_{\text{submitted}}$ follow the same qualitative pattern, with the largest difference being the amplitude of the peak error, which largely reflect the overall differences in forecast skill by model class (Fig. 6a). The timing of peak error varies slightly by model class with covariate peaking first (−9 days), followed by static (−6), DOY Mean (−5), persistence (−4) and dynamic (−3

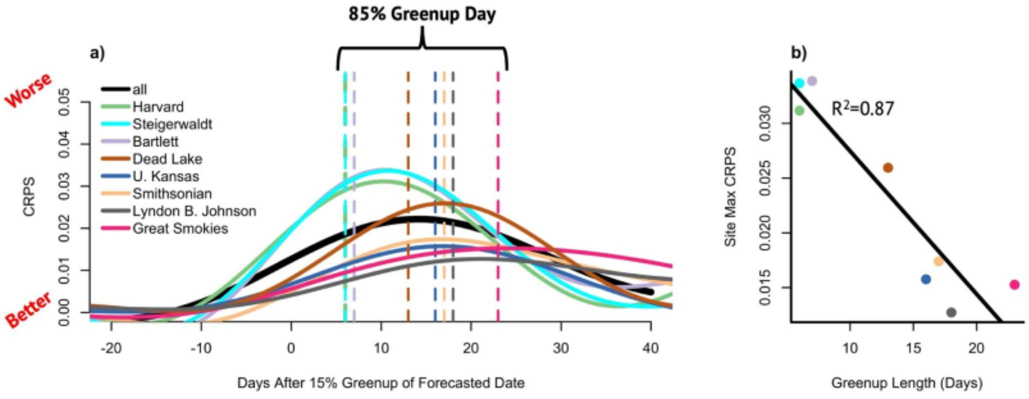


Fig. 5. a) Generalized Additive Model (GAM) response surfaces of Continuous Ranked Probability Score (CRPS) as a function of the predicted day relative to the date of 15 % greenup for each site based on GAM analyses shown in solid lines. The dotted vertical lines indicate the number of days after 15 % greenup that 85 % greenup occurred. Predictive power increased during the greenup period and for most sites peaked at or right after 85 % greenup. b) Site maximum CRPS *versus* greenup length (85 % greenup - 15 % greenup dates). Sites that greened up faster had worse predictability during greenup than sites that greened up slower.

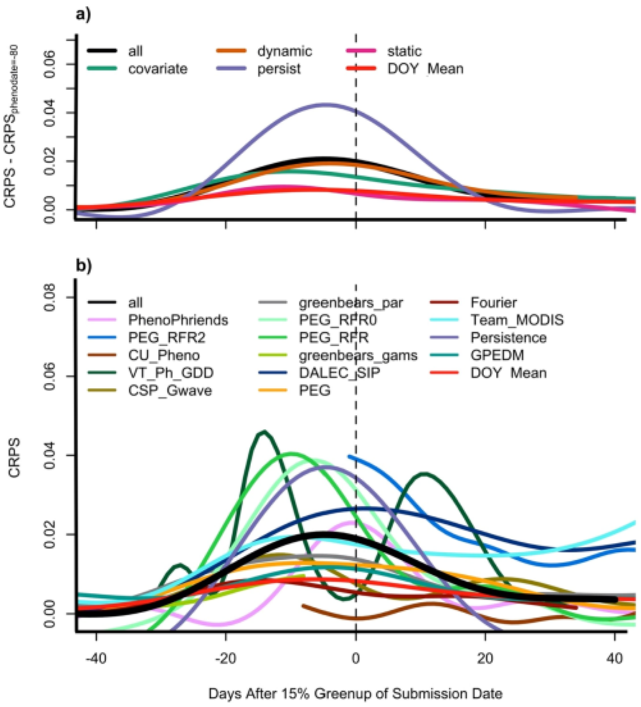


Fig. 6. a) Change in Continuous Ranked Probability Score (CRPS) in the Generalized Additive Model (GAM) response surfaces compared to the first day in the predicted time series (*i.e.*, phenodate$_{submitted}$ = −80) as a function of phenodate$_{submitted}$, defined as the submission date relative to the 15 % greenup date at each site. b) GAM predicted CRPS by model as a function of phenodate$_{submitted}$. In the overall GAM created from all forecasts, predictive error (CRPS), peaked four days before 15 % greenup.

days). Patterns by model are similar but with greater noise due to the variability in when teams first submitted forecasts and how often forecasts were submitted (Fig. 6b).

3.6. H4: predictability by model and model class

Contradicting H4, we found that the DOY Mean model class, not the dynamic models, overall had the highest predictive skill. Relative to the DOY Mean CRPS (ΔCRPS = 0), error was lowest for the covariate class (ΔCRPS=0.00168 ± 0.00012) followed by static (ΔCRPS=0.00245 ± 0.00011), dynamic (ΔCRPS=0.00697 ± 0.00009), and finally the persistence null (ΔCRPS=0.01176 ± 0.00010). Additionally in the overall GAM, only the model greenbears_par performed better than the DOY Mean null model ($\mu = -0.0005434, \sigma = 0.0001457, t = -3.729\ p = 0.000192$), while Fourier and EFI_U_P were not significantly different from the DOY Mean and all other models were significantly worse (Fig. 7). The mean standard deviation of submitted forecasts across the challenge period was significantly higher for the DOY Mean model than greenbears_par (0.00736 *versus* 0.00261; t = 102.9, degrees of freedom=35,303, p-value<2.2e−16). We also did not find a significant relationship between model effect in the overall GAM and the date of first submission (intercept: 6.042e−3; slope: −1.628e−5; R^2 = 0.01; F-statistic: 1.1693 on 1 and 15° of freedom; p-value: 0.6865). When grouped by model class instead of model, no model class significantly outperformed the DOY Mean null.

While only one of the models had higher predictive skill than DOY Mean across the entire Challenge period (greenbears_par), many models predicted G_{CC} on the transition dates better than the DOY Mean, sometimes 35 days out (Fig. S6). On average, for the 15 %, 50 %, and 85 % greenup transition dates, PEG, GPEDM, and greenbears_gams beat the DOY Mean model furthest out, respectively.

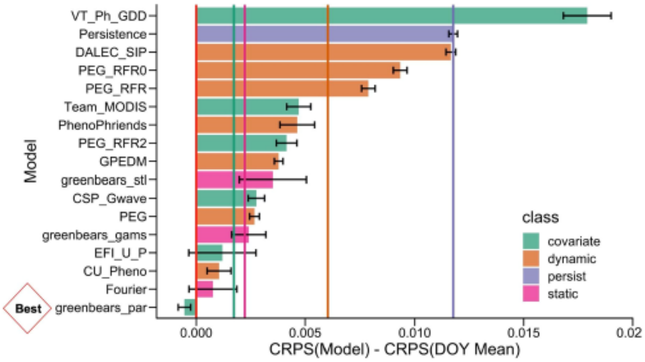


Fig. 7. Generalized Additive Model fixed effects expressing mean skill by model relative to the day of year (DOY) Mean null model. Models are ordered from highest error (top) to lowest error (bottom) and colored by model class. Negative values indicate the model outperformed the null across all forecasts. Vertical lines represent the effects for model class (red vertical line at 0 indicates DOY Mean). No model class significantly outperformed the DOY Mean null and greenbears_par was the only team to.

3.7. H5: predictability by site

Supporting H5, site effects (Fig. 8) exhibited a positive relationship with 50 % greenup DOY (slope = 8.847e−05, standard error = 3.063e-05, t = 2.888, p-$value$ = 0.0278) with an R^2 of 0.58, indicating that on average models were better at predicting sites that leafed out earlier than those that leafed out later. In terms of site-to-site differences in model performance, within the overall GAM all sites had significantly lower CRPS than the reference class (Bartlett) except Steigerwaldt.

In the GAM model that considered site × model interactions, 91 out of 144 interaction terms (67 %) were significant (Fig. S7). Interactions were least common for the models greenbears_stl (0), Fourier (0), greenbears (1), greenbears_gams (1), EFI_U_P (1), and PhenoPhriends (2). At the other extreme, all site interactions were significant for the models Team_MODIS, PEG_RFR0, PEG_RFR, persistence null, and DALEC_SIP, and seven out of eight sites were significant for PEG and greenbears_par.

4. Discussion

4.1. H1: skill and lead time

We observed that in general, and as expected, predictive skill of forecasts, as defined using CRPS that evaluates forecast distribution (rather than only the mean or median), increased as lead time decreased, which has been found with previous phenology forecasts (Taylor and White, 2020). That said, this overall pattern did not hold true for a couple specific cases. First, the DOY Mean null model showed no pattern with lead time, which is to be expected as this forecast is not updated based on new information and stays constant for each date regardless of when the forecast is created. For similar reasons teams using static models had, on average, the least increase in CRPS with lead time. The covariate model PEG_RFR2 showed the unexpected pattern of decreasing error with increased lead time, which was likely because the model's forecasts were only submitted at the end of the forecast period after most sites had already completed greenup. The covariate class's unexpected pattern of decreasing error with increased lead time likely occurred due to the models becoming overconfident at shorter lead times (*e.g.*, spread decreases more rapidly than bias with short lead times). This would also explain the initial behavior of the VT_Ph_GDD model of error declining as lead time increased only for the first week. At longer lead times (greater than approximately a week), the model either reduces bias or ensemble spread with decreasing lead times. Increases in CRPS, and thus decreased predictive skill, with shorter lead times has

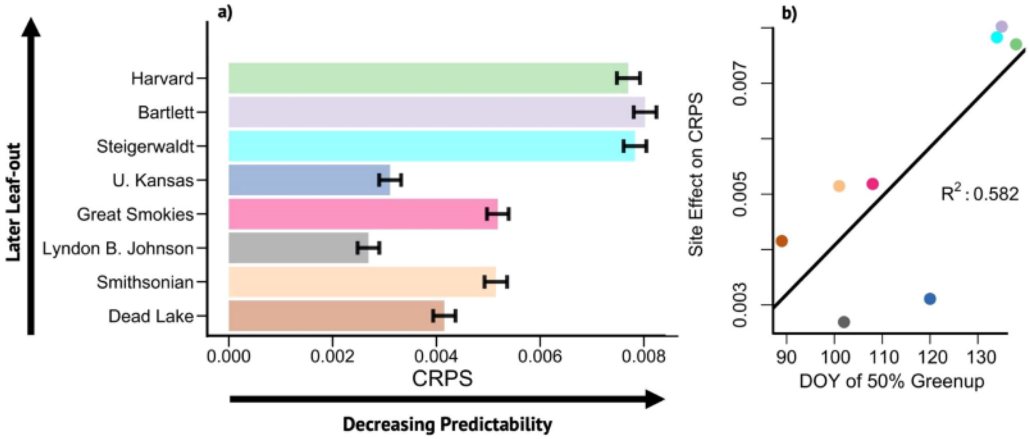


Fig. 8. a) Site effects on Continuous Ranked Probability Scores (CRPS) from the overall Generalized Additive Model ordered by date of leaf out from latest (top) to earliest (bottom). b) Linear regression of the site effects on CRPS from panel a *versus* the site's day of year (DOY) of 50 % greenup. Overall, sites that leafed out earlier had higher predictability.

been found elsewhere such as with forecasting streamflow totals (Schepen et al., 2016) and decadal hindcasts of global mean temperature (Smith et al., 2015). Our results emphasize that forecasters need to be wary of becoming overly confident with time.

4.2. H2: predictability of different parts of greenup

We hypothesized that the start of greenup would be the hardest part of the curve to forecast, but instead observed that in general, predictability decreased through the greenup period. This could be due to a variety of reasons. Firstly, the representation of budburst in the models could be better than the representation of leaf expansion. Budburst is typically more controlled by temperature and photoperiod in cold-deciduous plants (Chuine et al., 2013; Zohner et al., 2016), which was commonly used in the forecast models. In contrast, leaf and cell elongation is primarily controlled by water availability and turgor (Taiz and Zeiger, 2006), which was less common to include. Secondly, since the difference in greenness at 15 % greenup from dormancy is smaller than at later parts of greenup, if models are predicting greenup late the error on this date would be smaller but would grow over time until the forecast catches up with observations. Thirdly, the meteorological forecasts used could have been more inaccurate during later greenup instead of at budburst because of the impacts of vegetation properties leading to biases in meteorological forecasts (Xue et al., 1996). Focusing on model fits instead of near-term forecasts, Richardson et al. (2006) also found that later phenological stages are harder than budburst to predict. In contradiction, Klosterman et al. (2018) found that they were easier. Our results indicate that evaluating forecast and predictive skill at the end of the greenup period is as important as evaluating at budburst.

4.3. H3: predictability based on when submitted

The results support our hypothesis that in this focal year, selected sites, and submitted models, forecasts across all lead times submitted right before budburst have the lowest predictive power based on the selected metric of CRPS. This was expected because the 35-day forecasted period is long enough to typically include the full greenup curve in most deciduous broadleaf forests (Klosterman et al., 2018), and thus is harder to predict than the greenness during dormancy and peak greenness. Additionally, the error post-greenup was consistently higher than pre-greenup, which seems to be associated more with persistent biases across models in predicting peak summer greenness, than the potentially greater day-to-day variability in observations. During the summer, G_{CC} gradually decreases (Elmore et al., 2012; Klosterman et al., 2014), resulting in a less stable target compared to winter dormancy.

4.4. H4: skill of different model classes

We were surprised at how challenging it was to have higher skill than the DOY Mean model across the Challenge period because the DOY Mean model did not incorporate any covariates or current conditions. This difficulty could be partly attributed to the used G_{CC} index not being a perfect expression of phenology. Even with the PhenoCam Network processing (*e.g.*, fixing the white balance; Seyednasrollah et al., 2019), G_{CC} can still be affected by illumination and atmospheric conditions. Perhaps the DOY Mean model better accounted for these observation errors. The only model to have greater predictive skill, greenbears_par, relied on historical averages of the covariate data (DOY and photosynthetically active radiation), thus mimicking the historical average nature of the DOY Mean model. One reason it out-performed the DOY Mean model could be because the forecasts had lower uncertainties (mean standard deviation of 0.0026 *versus* 0.0074). This result reinforces just how important a historical means null model is for near-term forecasting in general, and for phenological forecasting in particular. While historical average models performed the best, they are likely less useful for predicting long-term changes to phenology as the climate warms, as they make the same prediction for every future year. It is also unclear, given one study year, how well the high performance of DOY Mean and greenbears_par would hold up across years that might be less "average." Furthermore while most models had lower predictive skill than the DOY Mean model across the Challenge period, we did observe that some models forecasted G_{CC} on the transition dates better than the DOY Mean model (Fig. S6), which is likely more important than across the entire Challenge period.

Historical means are often used as a null model in predicting specific transition dates, such as spring budburst, with mixed results of it outperforming other models. When comparing predictions of human-collected budburst timings in four species in Belgium, Fu et al. (2012) found that most models outperformed the historical null. However, many modeling studies, including those introducing a new model (*e.g.*, Elmendorf et al., 2019; García et al., 2019) and model comparison studies (*e.g.*, Asse et al., 2020; Melaas et al., 2016; Moon et al., 2021), do not include this as a null model. The difficulty in out-performing historical means predictions was also found by another theme of the EFI Ecological Forecasting Challenge (water temperature in lakes; Thomas et al., 2023b).

In contrast to H4, we found that other than the persistence model, the dynamic models (*i.e.*, ones that use the previous day's G_{CC} to make a forecast) performed in general worse than the other models, indicating that while dynamic phenology models still have potential to serve as mechanistic models and improve forecasts, the dynamic models used here for forecasts have likely not matured enough yet. We had expected that the low latency of the PhenoCam G_{CC} Data (overnight) would allow dynamic models to perform well because they could quickly incorporate the current state of the system before forecasting the next state. The conclusions here, however, are dependent on the specific models used in the forecasts as some dynamic models performed well, such as CU_Pheno. Furthermore, many common traditional phenology models (Chuine et al., 2013) could not be included because they forecast the timing of transition dates and not a timeseries of G_{CC} values. These commonly used models, though, are rarely dynamic models so including them would not have improved the performance of the dynamic model class.

4.5. H5: predictability differences between sites

In support of H5, we observed that the date of 50 % greenup and anomalies in the dates of 50 % and 85 % greenup each explained substantial variation in site predictability with sites that greened up later having lower predictability. Since we found no significant relationship between the date of starting to submit and model skill, this is likely not due to a non-random influx of late submitting models being worse than models that started submitting forecasts earlier. It is more likely a combination of ecological reasons. First, previous findings suggest that photoperiod is more of a dominant control of spring greenup in warmer climates in North America where temperature is more of a dominant control in colder climates (Moon et al., 2021) and in tree species found in lower latitudes (Zohner et al., 2016), which would explain why the DOY Mean performs better at sites in warmer climates that experience greenups earlier than the colder climates. Second, we found that the sites that had higher peak error (Steigerwaldt, Harvard, and Bartlett; Fig. 5) also greened up faster and occurred later than the other sites. This is similar to Klosterman et al. (2018) s finding that later springs greened-up faster, so it is possible that the models performed worse for these sites because of the faster than average greenup rates. Third, the National Phenology Network s published spring anomaly indices (Fig. S8) also suggest that the sites we found to be the hardest to forecast (Bartlett, Harvard, and Steigerwaldt) all had early springs in 2021. Including more years and a larger number of sites in future phenology forecasting challenges would help in assessing across-site patterns, as it is likely the small sample size (eight sites) limits the statistical power of such analyses. Similarly, with only one year of data it is hard to deconvolve across-site gradients in predictability from interannual variability, but these results generated hypotheses that we will use to approach future rounds.

4.6. Challenge evaluation

In addition to the scientific findings of the Challenge, we also observed numerous social aspects of the Challenge that were successful. We were successful at recruiting teams to submit forecasts for this first round despite a lack of a prize (*e.g.,* the 16,000 USD offered in Humphries et al., 2018) and limited prior experience across the phenology community in multi-team model intercomparisons. This is particularly noteworthy given that one of the decisions the design team made was for the NEON Phenology Forecasting Challenge to be based on forecasting greenness values at different days and not just the timing of transitions, which is typically emphasized in many classical phenological modeling approaches such as growing degree day thresholds. While this decision led to an underrepresentation of some of these classical modeling approaches, it led to innovative techniques, such as machine learning, and facilitated collaboration between computer science/machine learning experts and ecologists (*e.g.,* the PEG team models). Similarly, we had good participation by academic classes, which advanced training in ecological modeling and forecasting through hands-on experience. Finally, the infrastructure platform (Thomas et al., 2023a) that provided the data files of G_{CC} targets and GEFS meteorological forecasts, and received and displayed the forecasts, supported this real-time NEON Phenology Forecasting Challenge well and can support future challenges. Importantly, we succeeded in empowering many different teams of ecologists and data scientists to make genuine, probabilistic forecasts (*i.e.,* forecasts before data are collected)

While many aspects went well, there were some shortcomings to be improved upon in future rounds, especially aspects that would deconvolve the results. Firstly, not all teams submitted at all dates so there was a lack of consistency in the submissions presented challenges for intercomparison. Future rounds of the Challenge are set up to accept forecast submissions all year long (Thomas et al., 2023a). This is particularly important to encourage teams to consistently submit forecasts around the specific phenological events of greenup and senescence. Additionally, the small number of initial sites presented a challenge to understanding across-site patterns of predictability. To address this limitation, additional NEON sites have been added to the current and future rounds of the NEON Phenology Forecasting Challenge (increasing from eight to 47 sites and including other plant functional types). While these shortcomings will continue to be improved, intercomparison projects like this one lead to more creativity and ideas, which is exciting motivation as we continue the Challenge.

5. Conclusions

Here we presented the findings from the first round of a community spring greenup phenology forecast challenge. We found that in general predictability increases as lead time decreases (in support of H1); in this specific year and set of sites, predictive skill decreases at the later part of greenup (in contradiction to H2); forecasts submitted right before budburst had the lowest predictive skill (in support of H3); the DOY Mean null model is difficult to outperform across the entire greenup period (in contradiction to H4); and that sites that greened up later tended to be harder to predict (in support of H5). Our study emphasizes the importance of the historical means (or climatology) model as an important null model for ecological forecasting and improves our understanding of what affects the predictability of phenology. These findings should inform the focus of future forecasting and modeling efforts as we continue to investigate this important process as a broader community.

CRediT authorship contribution statement

Kathryn I. Wheeler: Conceptualization, Methodology, Formal analysis, Writing original draft, Writing review & editing, Visualization, Supervision. **Michael C. Dietze:** Conceptualization, Methodology, Formal analysis, Writing original draft, Writing review & editing, Visualization, Supervision, Funding acquisition. **David LeBauer:** Conceptualization, Methodology, Writing original draft, Writing review & editing. **Jody A. Peters:** Conceptualization, Methodology, Writing original draft, Writing review & editing, Project administration. **Andrew D. Richardson:** Conceptualization, Methodology, Writing original draft, Writing review & editing. **Arun A. Ross:** Conceptualization, Methodology, Writing original draft, Writing review & editing. **R. Quinn Thomas:** Conceptualization, Methodology, Software, Writing original draft, Writing review & editing, Visualization, Supervision, Funding acquisition. **Kai Zhu:** Conceptualization, Methodology, Writing original draft, Writing review & editing. **Uttam Bhat:** Methodology, Writing original draft. **Stephan Munch:** Methodology, Writing original draft. **Raphaela Floreani Buzbee:** Methodology, Writing original draft. **Min Chen:** Methodology, Writing original draft. **Benjamin Goldstein:** Methodology, Writing original draft. **Jessica Guo:** Methodology, Writing original draft. **Dalei Hao:** Methodology, Writing original draft. **Chris Jones:** Conceptualization, Methodology, Writing original draft. **Mira Kelly-Fair:** Methodology, Writing original draft. **Haoran Liu:** Methodology, Writing original draft. **Charlotte Malmborg:** Methodology, Writing original draft. **Naresh Neupane:** Methodology, Writing original draft. **Debasmita Pal:** Methodology, Writing original draft. **Vaughn Shirey:** Methodology, Writing original draft. **Yiluan Song:** Methodology, Writing original draft. **McKalee Steen:** Methodology, Writing original draft. **Eric A. Vance:** Methodology, Writing original draft. **Whitney M. Woelmer:** Methodology, Writing original draft. **Jacob H. Wynne:** Methodology, Writing original draft. **Luke Zachmann:** Methodology, Writing original draft.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Kathryn I Wheeler reports financial support was provided by National Oceanic and Atmospheric Administration. Andrew D. Richardson is on the Editorial Board of Agriculture and Forest Meteorology.

Data availability

The submitted forecasts with meta data and analysis code are archived on Zenodo (doi: 10.5281/zenodo.8200101). Code repositories for the individual models are given in the supplementary materials.

Acknowledgments

The EFI-RCN is supported by NSF (DEB-1926388) with computational support provided by NSF-funded Jetstream2 (OAC-2005506) and administrative support from an Alfred P. Sloan Foundation grant. NEON is a program sponsored by NSF and operated under cooperative agreement by Battelle. We thank our many collaborators, including site PIs and technicians, for their efforts in support of PhenoCam. The development of PhenoCam has been funded by the Northeastern States Research Cooperative, NSF s Macrosystems Biology program (awards EF-1065029 and EF-1702697), and DOE s Regional and Global Climate Modeling program (award DE-SC0016011). We acknowledge additional support from the US National Park Service Inventory and Monitoring Program and the USA National Phenology Network (grant number G10AP00129 from the United States Geological Survey), and from the USA National Phenology Network and North Central Climate Science Center (cooperative agreement number G16AC00224 from the United States Geological Survey). Additional funding, through the National Science Foundation s LTER program, has supported research at Harvard Forest (DEB-1237491) and Bartlett Experimental Forest (DEB-1114804). We also thank the USDA Forest Service Air Resource Management program and the National Park Service Air Resources program for contributing their camera imagery to the PhenoCam archive.

KIW, MCD, and CM were supported by the USA National Science Foundation (NSF) grant 1638577. KIW also acknowledges support under the NSF Graduate Research Fellowship grant 1247312 and a NOAA Climate and Global Change Postdoctoral Fellowship Program, administered by UCAR s Cooperative Programs for the Advancement of Earth System Science (CPAESS) under award #NA21OAR4310383. The PEG team (DL, JG, AAR, and DP) were supported by NSF awards 1940062 and 1940059. ADR acknowledges support from NSF awards 1702697, 1637685, and 1832210. RQT acknowledges support from NSF DEB award 1933016. GPEDM team (YS, UB, SBM, KZ) was supported by the University of California, Santa Cruz Hammett Fellowship, the Microsoft Azure AI for Earth Grant, and the NSF Award 2306198 (CAREER). HL and MC were supported by a McIntire Stennis grant (1027576) from the National Institute of Food and Agriculture (NIFA), United States Department of Agriculture (USDA). EAV acknowledges support from NSF grant 2022138. LZ was supported by NASA s AIST program (80NSSC17K0538), the Southwest Climate Adaptation Science Center (G17AC00247), and the Denver Zoological Foundation.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.agrformet.2023.109810.

References

Archetti, M., Richardson, A.D., O Keefe, J., Delpierre, N., 2013. Predicting climate change impacts on the amount and duration of autumn colors in a New England forest. PLoS One 8, e57373. https://doi.org/10.1371/journal.pone.0057373.

Asse, D., Randin, C.F., Bonhomme, M., Delestrade, A., Chuine, I., 2020. Process-based models outcompete correlative models in projecting spring phenology of trees in a future warmer climate. Agric. For. Meteorol. 285 286, 107931 https://doi.org/10.1016/j.agrformet.2020.107931.

Bloom, A.A., Williams, M., 2015. Constraining ecosystem carbon dynamics in a data-limited world: integrating ecological common sense in a model data fusion framework. Biogeosciences 12, 1299 1315. https://doi.org/10.5194/bg-12-1299-2015.

Boettiger, C., Thomas, R.Q., 2022. neon4cast: helper utilities for the EFI NEON forecast challenge. R package version 0.1.0.

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5 32. https://doi.org/10.1023/A:1010933404324.

Chen, M., Melaas, E.K., Gray, J.M., Friedl, M.A., Richardson, A.D., 2016. A new seasonal-deciduous spring phenology submodel in the Community Land Model 4.5: impacts on carbon and water cycling under future climate scenarios. Glob. Chang. Biol. 22, 3675 3688. https://doi.org/10.1111/gcb.13326.

Chuine, I., de Cortazar-Atauri, I.G., Kramer, K., Hanninen, H., 2013. Plant development models. In: Schwartz, M.D. (Ed.), Phenology: An Integrative Environmental Science. Springer, Netherlands, Dordrecht, pp. 275 293. https://doi.org/10.1007/978-94-007-6925-0_15.

Chuine, I., Regniere, J., 2017. Process-based models of phenology for plants and animals. Annu. Rev. Ecol. Evol. Syst. 48, 159 182. https://doi.org/10.1146/annurev-ecolsys-110316-022706.

Crimmins, T.M., 2020. The USA National Phenology Network: Big Ideas, Productivity, and Potential and Now, at Big Risk. The Bulletin of the Ecological Society of America 102 (1), e01802. https://doi.org/10.1002/bes2.1802.

Delpierre, N., Dufrene, E., Soudani, K., Ulrich, E., Cecchini, S., Boe, J., François, C., 2009. Modelling interannual and spatial variability of leaf senescence for three deciduous tree species in France. Agric. For. Meteorol. 149, 938 948. https://doi.org/10.1016/j.agrformet.2008.11.014.

Dietze, M.C., 2017. Ecological Forecasting. Princeton University Press, Princeton.

Dietze, M.C., Fox, A., Beck-Johnson, L.M., Betancourt, J.L., Hooten, M.B., Jarnevich, C.S., Keitt, T.H., Kenney, M.A., Laney, C.M., Larsen, L.G., Loescher, H.W., Lunch, C.K., Pijanowski, B.C., Randerson, J.T., Read, E.K., Tredennick, A.T., Vargas, R., Weathers, K.C., White, E.P., 2018. Iterative near-term ecological forecasting: needs, opportunities, and challenges. Proc. Natl. Acad. Sci. 115, 1424 1432. https://doi.org/10.1073/pnas.1710231115.

Dietze, M., Thomas, R.Q., Peters, J., Boettiger, C., Shiklomanov, A., 2023. A community convention for ecological forecasting: output files and metadata v1.0. Ecosphere 14, e4686. https://doi.org/10.1002/ecs2.4686.

Elmendorf, S.C., Crimmins, T.M., Gerst, K.L., Weltzin, J.F., 2019. Time to branch out? Application of hierarchical survival models in plant phenology. Agric. For. Meteorol. 279, 107694 https://doi.org/10.1016/j.agrformet.2019.107694.

Elmore, A.J., Guinn, S.M., Minsley, B.J., Richardson, A.D., 2012. Landscape controls on the timing of spring, autumn, and growing season length in mid-Atlantic forests. Glob Chang Biol 18, 656 674. https://doi.org/10.1111/j.1365-2486.2011.02521.x.

Filippa, G., Cremonese, E., Migliavacca, M., Galvagno, M., Folker, M., Richardson, A.D., Tomelleri, E., 2020. phenopix: process digital images of a vegetation cover.

Fu, Y.H., Campioli, M., Van Oijen, M., Deckmyn, G., Janssens, I.A., 2012. Bayesian comparison of six different temperature-based budburst models for four temperate tree species. Ecol. Modell. 230, 92 100. https://doi.org/10.1016/j.ecolmodel.2012.01.010.

García, M.A., Moutahir, H., Casady, G.M., Bautista, S., Rodríguez, F., 2019. Using hidden markov models for land surface phenology: an evaluation across a range of land cover types in southeast Spain. Remote Sens. (Basel) 11, 507. https://doi.org/10.3390/rs11050507.

Gerst, K.L., Crimmins, T.M., Posthumus, E., Marsh, R.L., Switzer, J., Wallace, C., 2021. The USA national phenology network s Buffelgrass green-up forecast map products. Ecol. Solut. Evid. 2, e12109 https://doi.org/10.1002/2688-8319.12109.

Humphries, G.R.W., Che-Castaldo, C., Bull, P.J., Lipstein, G., Ravia, A., Carrion, B., Bolton, T., Ganguly, A., Lynch, H.J., 2018. Predicting the future is hard and other lessons from a population time series data science competition. Ecol. Inform. 48, 1 11. https://doi.org/10.1016/j.ecoinf.2018.07.004.

Hyndman, R.J., Khandakar, Y., 2008. Automatic time series forecasting: the forecast package for R. J. Stat. Softw. 27, 1 22. https://doi.org/10.18637/jss.v027.i03.

Johansson, M.A., Apfeldorf, K.M., Dobson, S., Devita, J., Buczak, A.L., Baugher, B., Moniz, L.J., Bagley, T., Babin, S.M., Guven, E., Yamana, T.K., Shaman, J., Moschou, T., Lothian, N., Lane, A., Osborne, G., Jiang, G., Brooks, L.C., Farrow, D.C., Hyun, S., Tibshirani, R.J., Rosenfeld, R., Lessler, J., Reich, N.G., Cummings, D.A.T., Lauer, S.A., Moore, S.M., Clapham, H.E., Lowe, R., Bailey, T.C., García-Díez, M., Carvalho, M.S., Rodo, X., Sardar, T., Paul, R., Ray, E.L., Sakrejda, K., Brown, A.C., Meng, X., Osoba, O., Vardavas, R., Manheim, D., Moore, M., Rao, D.M., Porco, T.C., Ackley, S., Liu, F., Worden, L., Convertino, M., Liu, Y., Reddy, A., Ortiz, E., Rivero, J., Brito, H., Juarrero, A., Johnson, L.R., Gramacy, R.B., Cohen, J.M., Mordecai, E.A., Murdock, C.C., Rohr, J.R., Ryan, S.J., Stewart-Ibarra, A.M., Weikel, D.P., Jutla, A., Khan, R., Poultney, M., Colwell, R.R., Rivera-García, B., Barker, C.M., Bell, J.E., Biggerstaff, M., Swerdlow, D., Mier-y-Teran-Romero, L., Forshey, B.M., Trtanj, J., Asher, J., Clay, M., Margolis, H.S., Hebbeler, A.M., George, D., Chretien, J.P., 2019. An open challenge to advance probabilistic forecasting for dengue epidemics. Proc. Natl. Acad. Sci. 116, 24268 24274. https://doi.org/10.1073/pnas.1909865116.

Jordan, A., Krüger, F., Lerch, S., 2019. Evaluating probabilistic forecasts with scoringRules. J. Stat. Softw. 90, 1 37. https://doi.org/10.18637/jss.v090.i12.

Keenan, T.F., Richardson, A.D., 2015. The timing of autumn senescence is affected by the timing of spring phenology: implications for predictive models. Glob. Chang Biol. 21, 2634 2641. https://doi.org/10.1111/gcb.12890.

Klosterman, S., Hufkens, K., Richardson, A.D., 2018. Later springs green-up faster: the relation between onset and completion of green-up in deciduous forests of North America. Int. J. Biometeorol. 62, 1645 1655. https://doi.org/10.1007/s00484-018-1564-9.

Klosterman, S.T., Hufkens, K., Gray, J.M., Melaas, E., Sonnentag, O., Lavine, I., Mitchell, L., Norman, R., Friedl, M.A., Richardson, A.D., 2014. Evaluating remote sensing of deciduous forest phenology at multiple spatial scales using PhenoCam imagery. Biogeosciences 11, 4305 4320. https://doi.org/10.5194/bg-11-4305-2014.

Lebourgeois, F., Pierrat, J.C., Perez, V., Piedallu, C., Cecchini, S., Ulrich, E., 2010. Simulating phenological shifts in French temperate forests under two climatic change scenarios and four driving global circulation models. Int. J. Biometeorol. 54, 563 581. https://doi.org/10.1007/s00484-010-0305-5.

Li, W., Guan, H., Zhu, Y., Zhou, X., Fu, B., Hou, D., Sinsky, E., Xue, X., 2019. Prediction Skill of the MJO, NAO and PNA in the NCEP FV3-GEFS 35-day Experiments. Science and Technology Infusion Climate Bulletin. NOAA s National Weather Service, Durham, NA, p. 4.

Melaas, E.K., Friedl, M.A., Richardson, A.D., 2016. Multiscale modeling of spring phenology across deciduous forests in the Eastern United States. Glob. Chang. Biol. 22, 792 805. https://doi.org/10.1111/gcb.13122.

Milliman, T., Seyednasrollah, B., Young, A.M., Hufkens, K., Friedl, M.A., Frolking, S., Richardson, A.D., Abraha, M., Allen, D.W., Apple, M., Arain, M.A., Baker, J., Baker, J.M., Bernacchi, C.J., Bhattacharjee, J., Blanken, P., Bosch, D.D., Boughton, R., Boughton, E.H., Brown, R.F., Browning, D.M., Brunsell, N., Burns, S.P., Cavagna, M., Chu, H., Clark, P.E., Conrad, B.J., Cremonese, E., Debinski, D., Desai, A.R., Diaz-Delgado, R., Duchesne, L., Dunn, A.L., Eissenstat, D.M., El-Madany, T., Ellum, D.S.S., Ernest, S.M., Esposito, A., Fenstermaker, L., Flanagan, L. B., Forsythe, B., Gallagher, J., Gianelle, D., Griffis, T., Groffman, P., Gu, L., Guillemot, J., Halpin, M., Hanson, P.J., Hemming, D., Hove, A.A., Humphreys, E.R., Jaimes-Hernandez, A., Jaradat, A.A., Johnson, J., Keel, E., Kelly, V.R., Kirchner, J. W., Kirchner, P.B., Knapp, M., Krassovski, M., Langvall, O., Lanthier, G., Maire, G.l., Magliulo, E., Martin, T.A., McNeil, B., Meyer, G.A., Migliavacca, M., Mohanty, B.P., Moore, C.E., Mudd, R., Munger, J.W., Murrell, Z.E., Nesic, Z., Neufeld, H.S., Oechel, W., Oishi, A.C., Oswald, W.W., Perkins, T.D., Reba, M.L., Rundquist, B., Runkle, B.R., Russell, E.S., Sadler, E.J., Saha, A., Saliendra, N.Z., Schmalbeck, L., Schwartz, M.D., Scott, R.L., Smith, E.M., Sonnentag, O., Stoy, P., Strachan, S., Suvocarev, K., Thom, J.E., Thomas, R.Q., Van den berg, A.K., Vargas, R., Vogel, C.S., Walker, J.J., Webb, N., Wetzel, P., Weyers, S., Whipple, A.V., Whitham, T.G., Wohlfahrt, G., Wood, J.D., Yang, J., Yang, X., Yenni, G., Zhang, Y., Zhang, Q., Zona, D., Baldocchi, D., Verfaillie, J.. PhenoCam dataset v2.0: digital camera imagery from the PhenoCam network, 2000 2018. ORNL DAAC, Oak Ridge, Tennessee, USA. https://doi.org/10.3334/ORNLDAAC/1689.

Moon, M., Seyednasrollah, B., Richardson, A.D., Friedl, M.A., 2021. Using time series of MODIS land surface phenology to model temperature and photoperiod controls on spring greenup in North American deciduous forests. Remote Sens. Environ. 260, 112466 https://doi.org/10.1016/j.rse.2021.112466.

Morisette, J.T., Richardson, A.D., Knapp, A.K., Fisher, J.I., Graham, E.A., Abatzoglou, J., Wilson, B.E., Breshears, D.D., Henebry, G.M., Hanes, J.M., Liang, L., 2009. Tracking the rhythm of the seasons in the face of global change: phenological research in the 21st century. Front. Ecol. Environ. 7, 253 260. https://doi.org/10.1890/070217.

Munch, S.B., Poynor, V., Arriaza, J.L., 2017. Circumventing structural uncertainty: a Bayesian perspective on nonlinear forecasting for ecology. Ecol. Complex., Uncertain. Ecol. 32, 134 143. https://doi.org/10.1016/j.ecocom.2016.08.006.

Neupane, N., Peruzzi, M., Arab, A., Mayor, S.J., Withey, J.C., Ries, L., Finley, A.O., 2022. A novel model to accurately predict continental-scale timing of forest green-up. Int. J. Appl. Earth Observ. Geoinform. 108, 102747 https://doi.org/10.1016/j.jag.2022.102747.

Parmesan, C., Yohe, G., 2003. A globally coherent fingerprint of climate change impacts across natural systems. Nature 421, 37 42. https://doi.org/10.1038/nature01286.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E., 2011. Scikit-learn: machine learning in Python. J. Mach. Learn. Res. 12, 2825 2830.

Piao, S., Liu, Q., Chen, A., Janssens, I.A., Fu, Y., Dai, J., Liu, L., Lian, X., Shen, M., Zhu, X., 2019. Plant phenology and global climate change: current progresses and challenges. Glob. Chang Biol. 25, 1922 1940. https://doi.org/10.1111/gcb.14619.

R Core Team, 2022. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

Richardson, A.D., Bailey, A.S., Denny, E.G., Martin, C.W., O keefe, J., 2006. Phenology of a northern hardwood forest canopy. Glob. Chang Biol. 12, 1174 1188. https://doi.org/10.1111/j.1365-2486.2006.01164.x.

Richardson, A.D., Hufkens, K., Milliman, T., Aubrecht, D.M., Chen, M., Gray, J.M., Johnston, M.R., Keenan, T.F., Klosterman, S.T., Kosmala, M., Melaas, E.K., Friedl, M. A., Frolking, S., 2018. Tracking vegetation phenology across diverse North American biomes using PhenoCam imagery. Sci. Data 5, 180028. https://doi.org/10.1038/sdata.2018.28.

Richardson, A.D., Keenan, T.F., Migliavacca, M., Ryu, Y., Sonnentag, O., Toomey, M., 2013. Climate change, phenology, and phenological control of vegetation feedbacks to the climate system. Agric. For. Meteorol. 169, 156 173. https://doi.org/10.1016/j.agrformet.2012.09.012.

Schepen, A., Zhao, T., Wang, Q.J., Zhou, S., Feikema, P., 2016. Optimising seasonal streamflow forecast lead time for operational decision making in Australia. Hydrol. Earth Syst. Sci. 20, 4117 4128. https://doi.org/10.5194/hess-20-4117-2016.

Seyednasrollah, B., Young, A.M., Hufkens, K., Milliman, T., Friedl, M.A., Frolking, S., Richardson, A.D., 2019. Tracking vegetation phenology across diverse biomes using Version 2.0 of the PhenoCam Dataset. Sci. Data 6, 222. https://doi.org/10.1038/s41597-019-0229-9.

Smith, L.A., Suckling, E.B., Thompson, E.L., Maynard, T., Du, H., 2015. Towards improving the framework for probabilistic forecast evaluation. Clim. Change 132, 31 45. https://doi.org/10.1007/s10584-015-1430-2.

Sonnentag, O., Hufkens, K., Teshera-Sterne, C., Young, A.M., Friedl, M., Braswell, B.H., Milliman, T., O Keefe, J., Richardson, A.D., 2012. Digital repeat photography for phenological research in forest ecosystems. Agric. For. Meteorol. 152, 159 177. https://doi.org/10.1016/j.agrformet.2011.09.009.

Stockli, R., Rutishauser, T., Dragoni, D., O Keefe, J., Thornton, P.E., Jolly, M., Lu, L., Denning, A.S., 2008. Remote sensing data assimilation for a prognostic phenology model: data assimilation and phenology modeling. J. Geophys. Res. 113 https://doi.org/10.1029/2008JG000781.

Taiz, L., Zeiger, E., 2006. Plant Physiology, 4th. ed. Sinauer Associates, Inc., Sunderland, Massachussetts.

Taylor, S.D., White, E.P., 2020. Automated data-intensive forecasting of plant phenology throughout the United States. Ecol. Appl. 30, e02025 https://doi.org/10.1002/eap.2025.

Thomas, R.Q., Boettiger, C., Carey, C.C., Dietze, M.C., Johnson, L.R., Kenney, M.A., Mclachlan, J.S., Peters, J.A., Sokol, E.R., Weltzin, J.F., Willson, A., Woelmer, W.M., Challenge Contributors, 2023a. The NEON Ecological Forecasting Challenge. Frontiers in Ecology and Environment 21, 112 113. https://doi.org/10.1002/fee.2616.

Thomas, R.Q., McClure, R.P., Moore, T.N., Woelmer, W.M., Boettiger, C., Figueiredo, R. J., Hensley, R.T., Carey, C.C., 2023b. Near-term forecasts of NEON lakes reveal gradients of environmental predictability across the U.S. Frontiers in Ecology and Environment 21, 220 226. https://doi.org/10.1002/fee.2623.

Thornton, P.E., THORNTON, M.M., MAYER, B.W., WEI, Y., DEVARAKONDA, R., VOSE, R.S., COOK, R.B., 2017. Daymet: Daily Surface Weather Data On a 1-km Grid for North America, Version 3. ORNL Distributed Active Archive Center. https://doi.org/10.3334/ORNLDAAC/1328.

Toomey, M., Friedl, M.A., Frolking, S., Hufkens, K., Klosterman, S., Sonnentag, O., Baldocchi, D.D., Bernacchi, C.J., Biraud, S.C., Bohrer, G., Brzostek, E., Burns, S.P., Coursolle, C., Hollinger, D.Y., Margolis, H.A., McCaughey, H., Monson, R.K., Munger, J.W., Pallardy, S., Phillips, R.P., Torn, M.S., Wharton, S., Zeri, M., Richardson, A.D., 2015. Greenness indices from digital cameras predict the timing and seasonal dynamics of canopy-scale photosynthesis. Ecol. Appli. 25, 99 115. https://doi.org/10.1890/14-0005.1.

USA National Phenology Network, 2017. Historical annual spring indices anomaly (2016-Previous Year), First Leaf - Spring Index, Year: 2021. Region: 49.9375,-66.4791667,24.0625,-125.0208333. 10.5066/F7XD0ZRK.

Viskari, T., Hardiman, B., Desai, A.R., Dietze, M.C., 2015. Model-data assimilation of multiple phenological observations to constrain and predict leaf area index. Ecol. Appl. 25, 546 558. https://doi.org/10.1890/14-0497.1.

Wood, S., 2022. mgcv: mixed GAM computation vehicle with automatic smoothness estimation.

Wood, S.N., 2017. Generalized Additive Models: an Introduction with R (2nd edition).

Wu, S., Zeng, Y., Hao, D., Liu, Q., Li, J., Chen, X., Asrar, G.R., Yin, G., Wen, J., Yang, B., Zhu, P., Chen, M., 2021. Quantifying leaf optical properties with spectral invariants theory. Remote Sens. Environ. 253, 112131 https://doi.org/10.1016/j.rse.2020.112131.

Xie, Y., Wang, X., Wilson, A.M., Silander, J.A., 2018. Predicting autumn phenology: how deciduous tree species respond to weather stressors. Agric. For. Meteorol. 250 251, 127 137. https://doi.org/10.1016/j.agrformet.2017.12.259.

Xu, X., Tang, Y., Qu, Y., Zhou, Z., Hu, J., 2021. Global vegetation photosynthetic phenology products based on MODIS vegetation greenness and temperature: modeling and evaluation. Remote Sens. 13, 5080. https://doi.org/10.3390/rs13245080.

Xue, Y., Fennessy, M.J., Sellers, P.J., 1996. Impact of vegetation properties on U.S. summer weather prediction. J. Geophys. Res.: Atmosp. 101, 7419 7430. https://doi.org/10.1029/95JD02169.

Yang, X., Tang, J., Mustard, J.F., Wu, J., Zhao, K., Serbin, S., Lee, J.E., 2016. Seasonal variability of multiple leaf traits captured by leaf spectroscopy at two temperate deciduous forests. Remote Sens. Environ. 179, 1 12. https://doi.org/10.1016/j.rse.2016.03.026.

Zeng, Y., Xu, B., Yin, G., Wu, S., Hu, G., Yan, K., Yang, B., Song, W., Li, J., 2018. Spectral invariant provides a practical modeling approach for future biophysical variable estimations. Remote Sens. 10, 1508. https://doi.org/10.3390/rs10101508.

Zohner, C.M., Benito, B.M., Svenning, J.C., Renner, S.S., 2016. Day length unlikely to constrain climate-driven shifts in leaf-out times of northern woody plants. Nat. Clim. Change 6, 1120 1123. https://doi.org/10.1038/nclimate3138.