Supplementary Figures

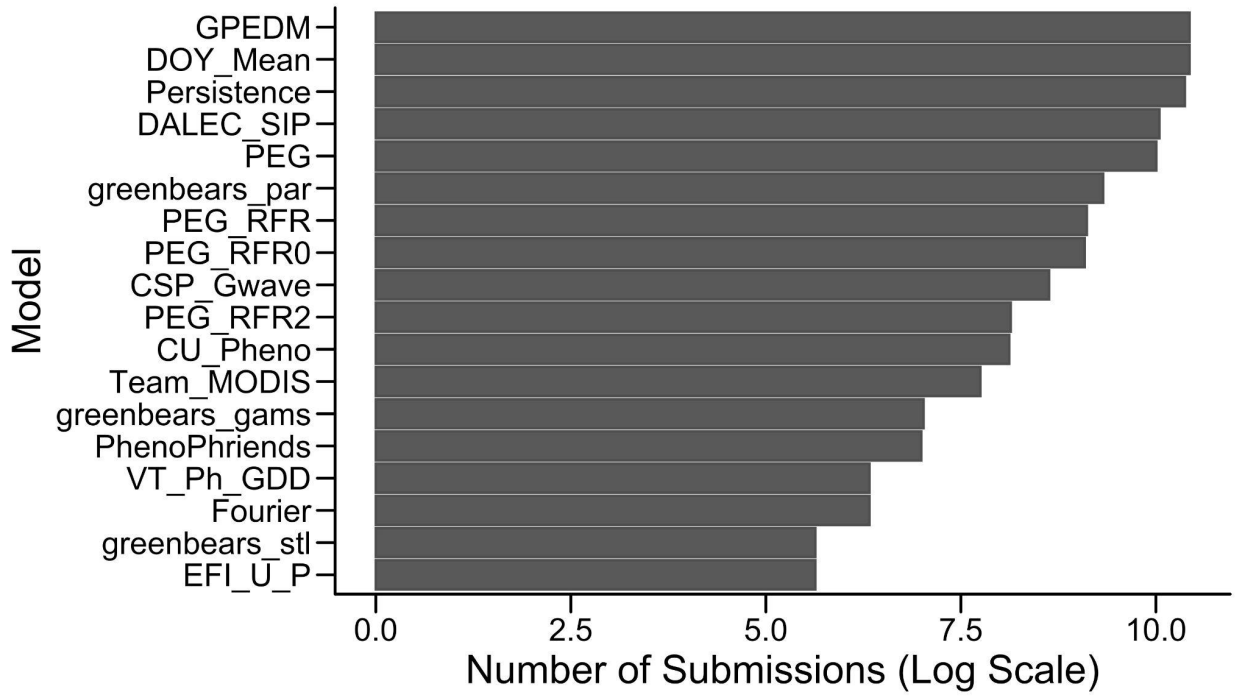

Figure S1. Number of total forecast submissions on a log scale per model throughout the Challenge period. The number of forecasts varied greatly between models.

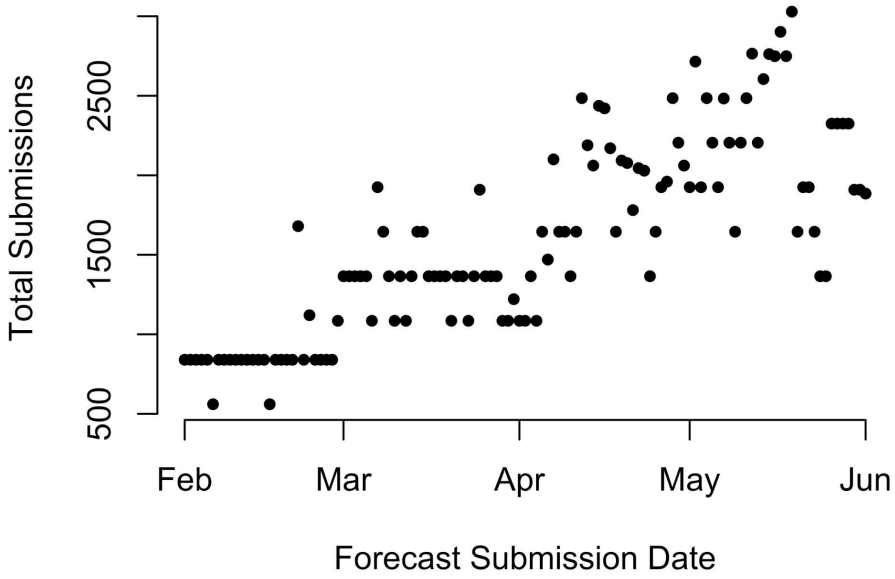

Figure S2. Count of forecasts for each submission date. In general, submissions increased throughout the Challenge until mid-May.

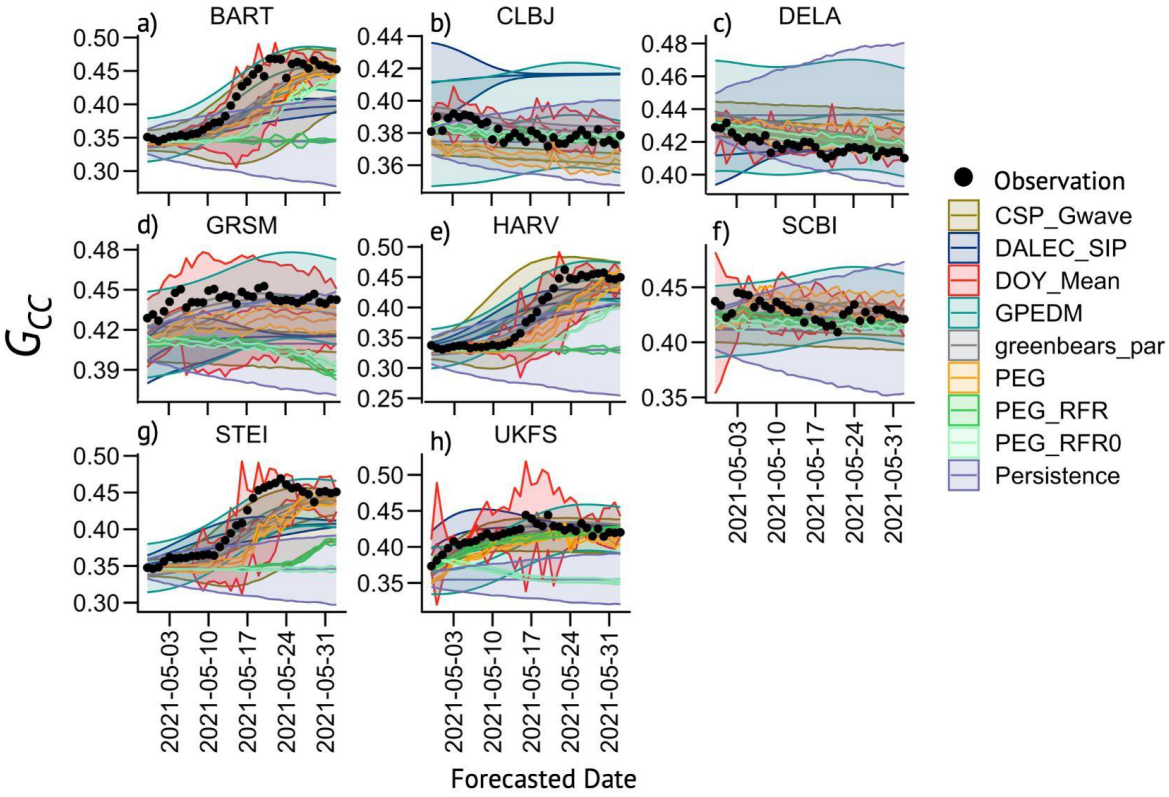

Figure S3. The forecasts of greenness (G_{CC}) that were submitted on 28 April 2021 for the next 35 days at the eight sites. For site descriptions and name abbreviations see Table 1.

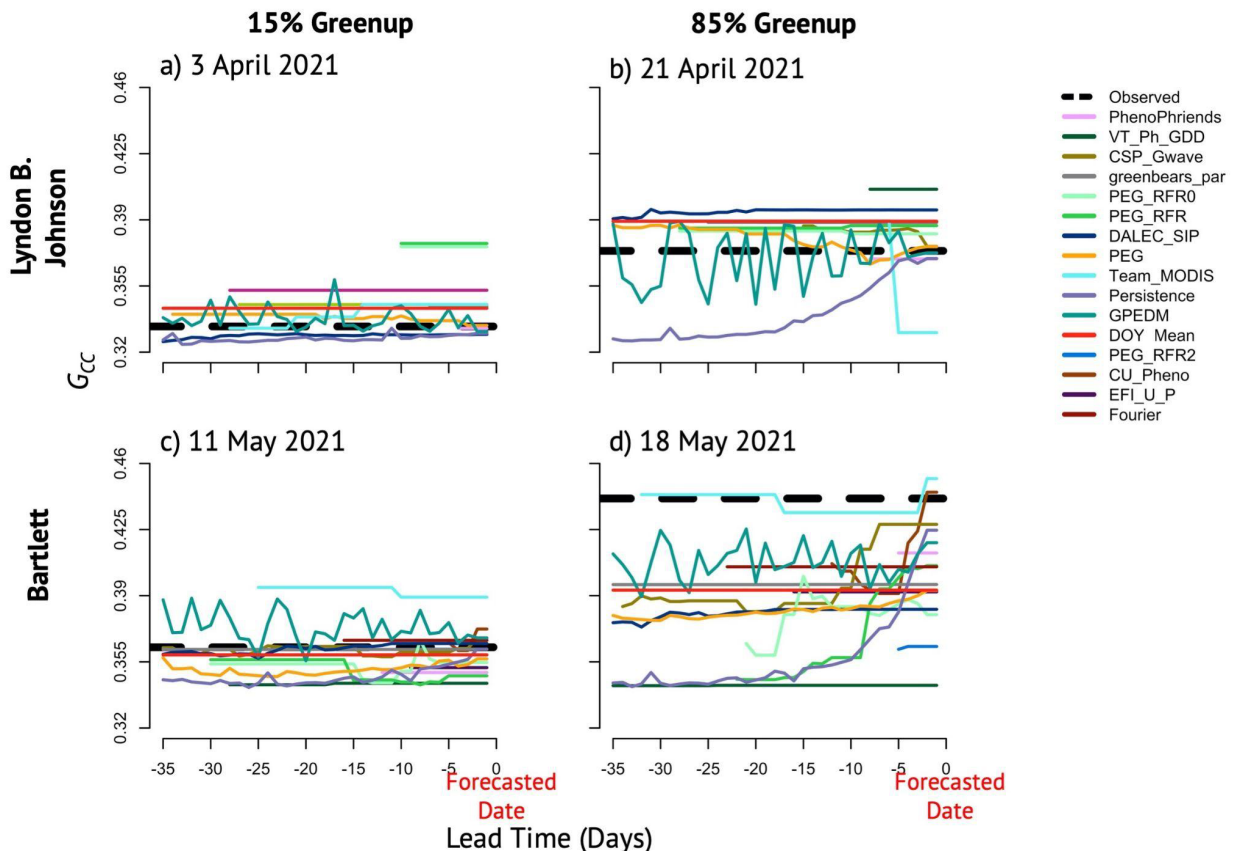

Figure S4. Examples of how teams' forecasts of greenness (G_{CC}) change as transition dates (15% and 85% greenup in left and right columns, respectively) approach and lead time decreases for Lyndon B. Johnson National Grassland on the top row and Bartlett Experimental Forest on the bottom row. (a) forecasts for Lyndon B. Johnson on 3 April 2021, (b) forecasts for Lyndon B. Johnson on 21 April 2021, (c) forecasts for Bartlett on 11 May 2021, and (d) forecasts for Bartlett on 18 May 2021. Some models' forecasts approached the observed G_{CC} as lead time decreased and some did not.

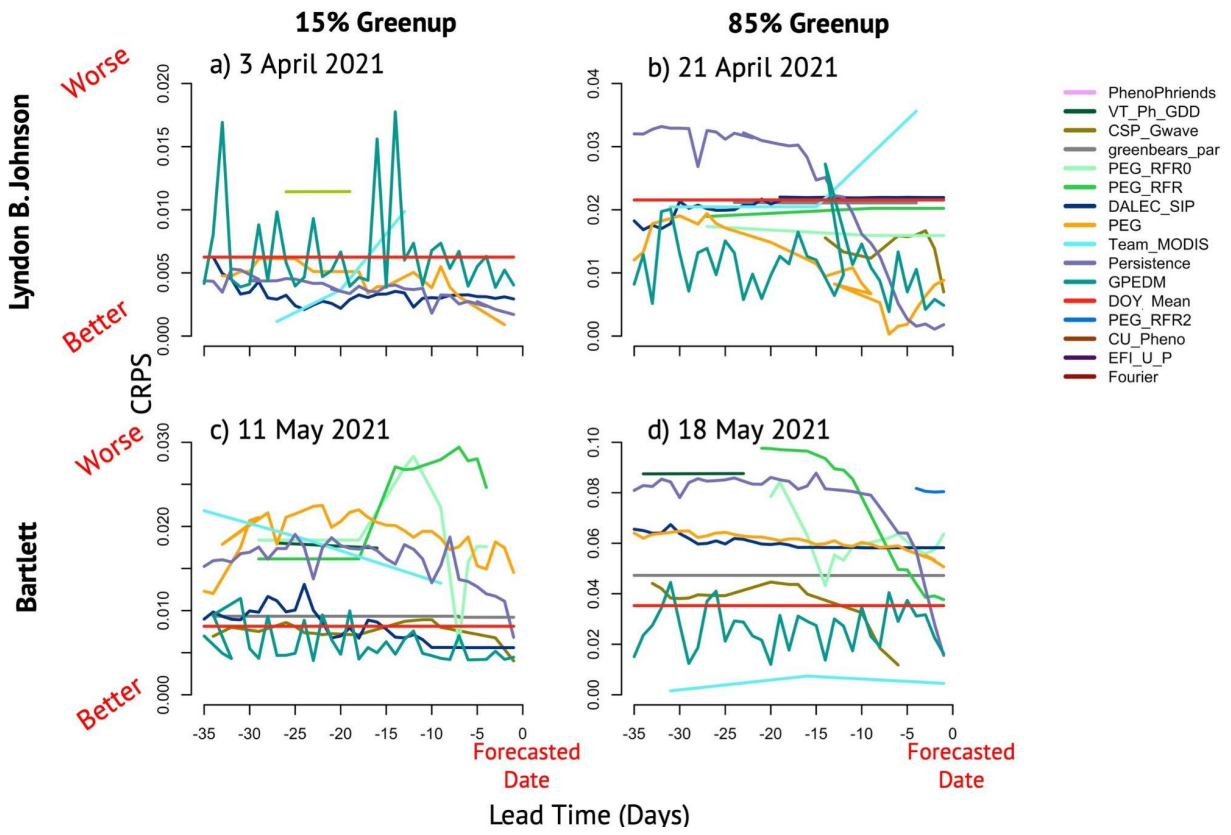

Figure S5. Examples of how the Continuous Ranked Probability Scores (CRPS) of teams' forecasts change as transition dates (15% and 85% greenup in left and right columns, respectively) approach and lead time decreases for Lyndon B. Johnson National Grassland on the top row and Bartlett Experimental Forest on the bottom row. (a) forecasts for Lyndon B. Johnson on 3 April 2021, (b) forecasts for Lyndon B. Johnson on 21 April 2021, (c) forecasts for Bartlett on 11 May 2021, and (d) forecasts for Bartlett on 18 May 2021. Forecast predictive skill (lower CRPS) was higher for Lyndon B. Johnson than for Bartlett.

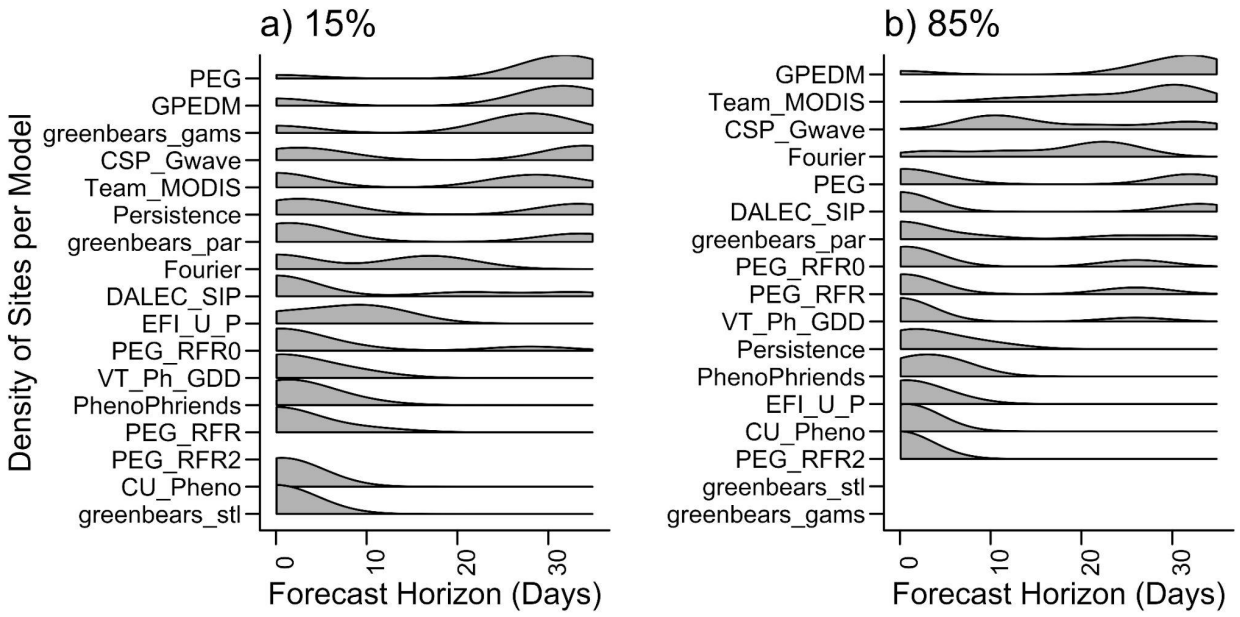

Figure S6. Forecast horizon, or number of days before the transition dates that each forecast model did better at forecasting greenness (G_{CC}) than day of year mean model across the range of all sites based on Continuous Ranked Probability Score values for the dates of 15% (a) and 85% (b) greenup. In each panel, the models are ordered by decreasing average forecast horizon. Empty rows indicate that the team did not forecast G_{CC} on the relevant dates for any sites. On average, PEG and GPEDM models forecasted G_{CC} on the 15% and 85% transition dates, respectively, better than the day of year mean model the furthest in advance.

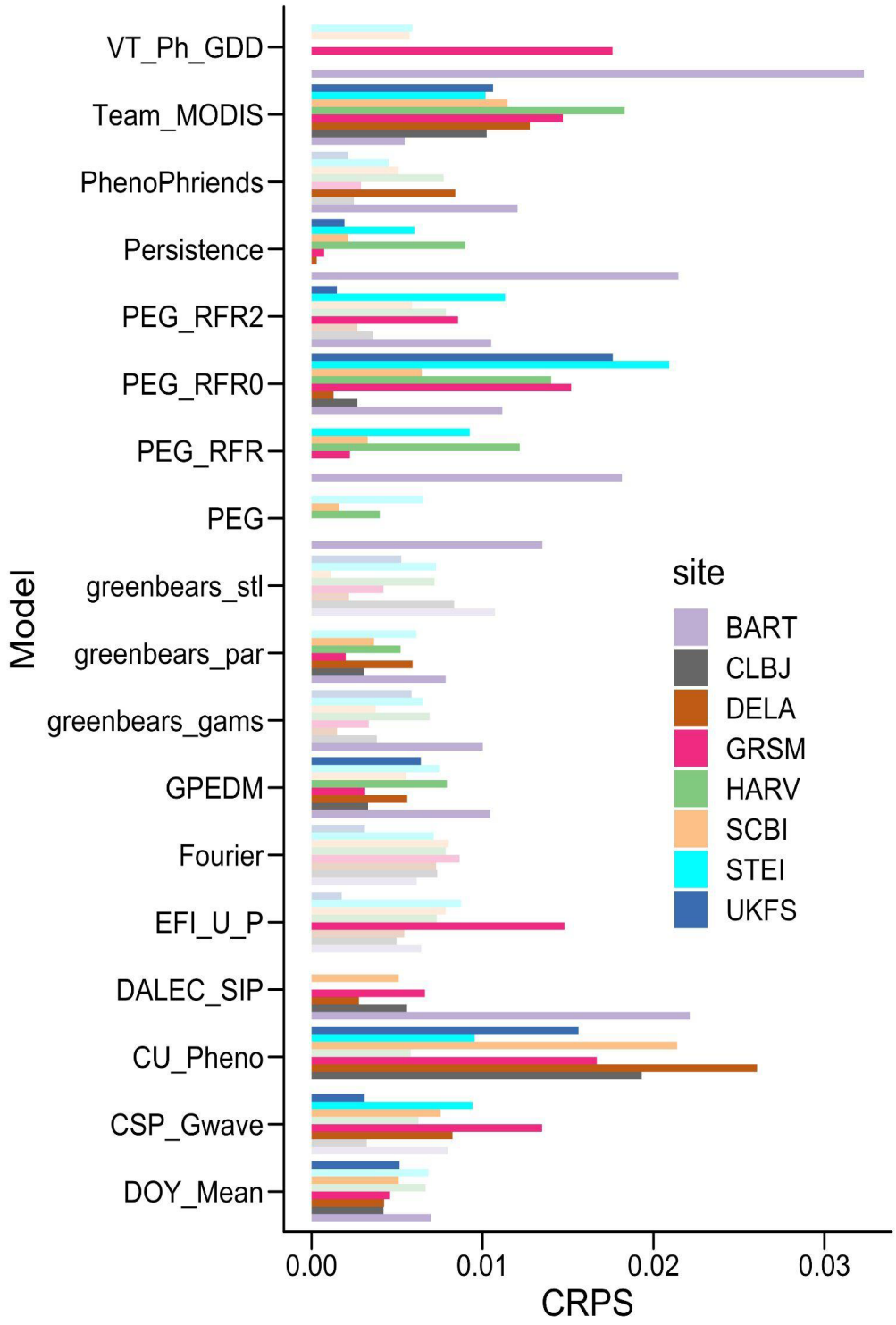

Figure S7: Generalized Additive Model fixed effects for each model x site combination. Nonsignificant interaction effects are indicated using transparency.

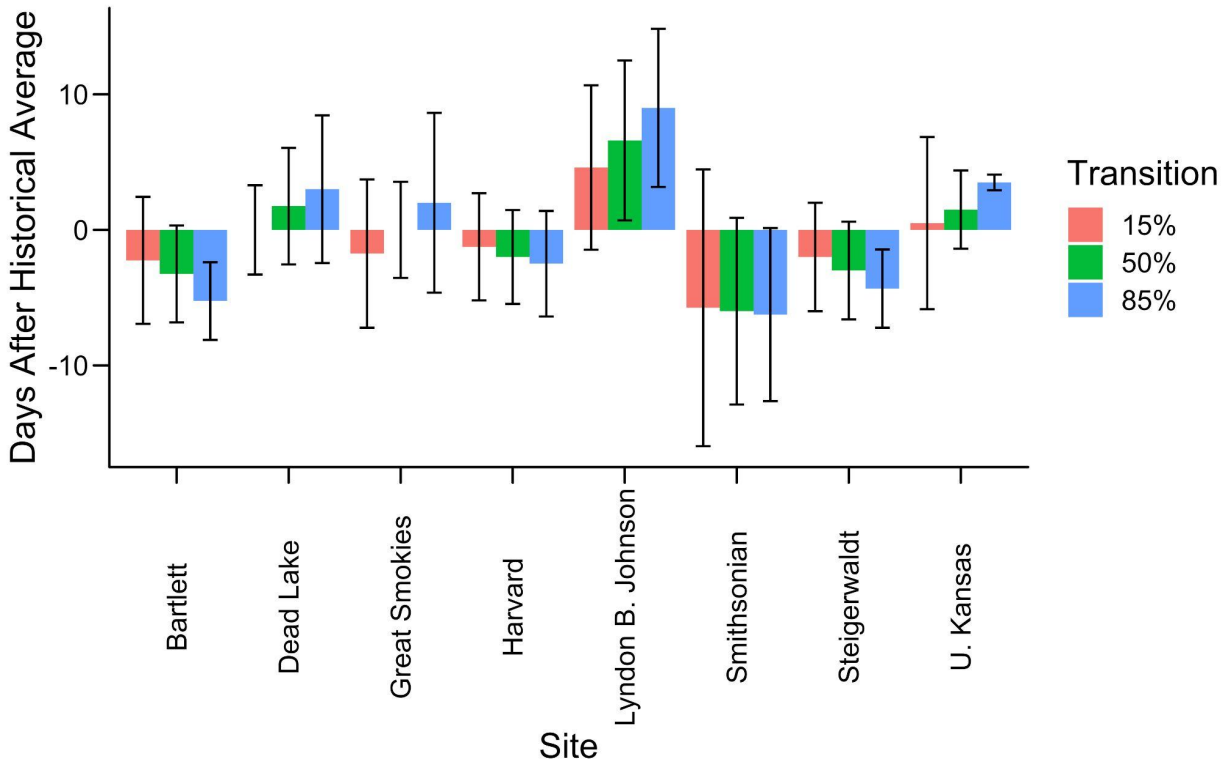

Figure S8. The number of days anomaly in the 2021 calculated dates of 15%, 50%, and 85% greenup compared to the average in the available historical PhenoCam data for each site. Positive values indicate that the year was later than average and negative values indicate earlier. The standard deviations in the historical data (two – four years for each site) are given in error bars.

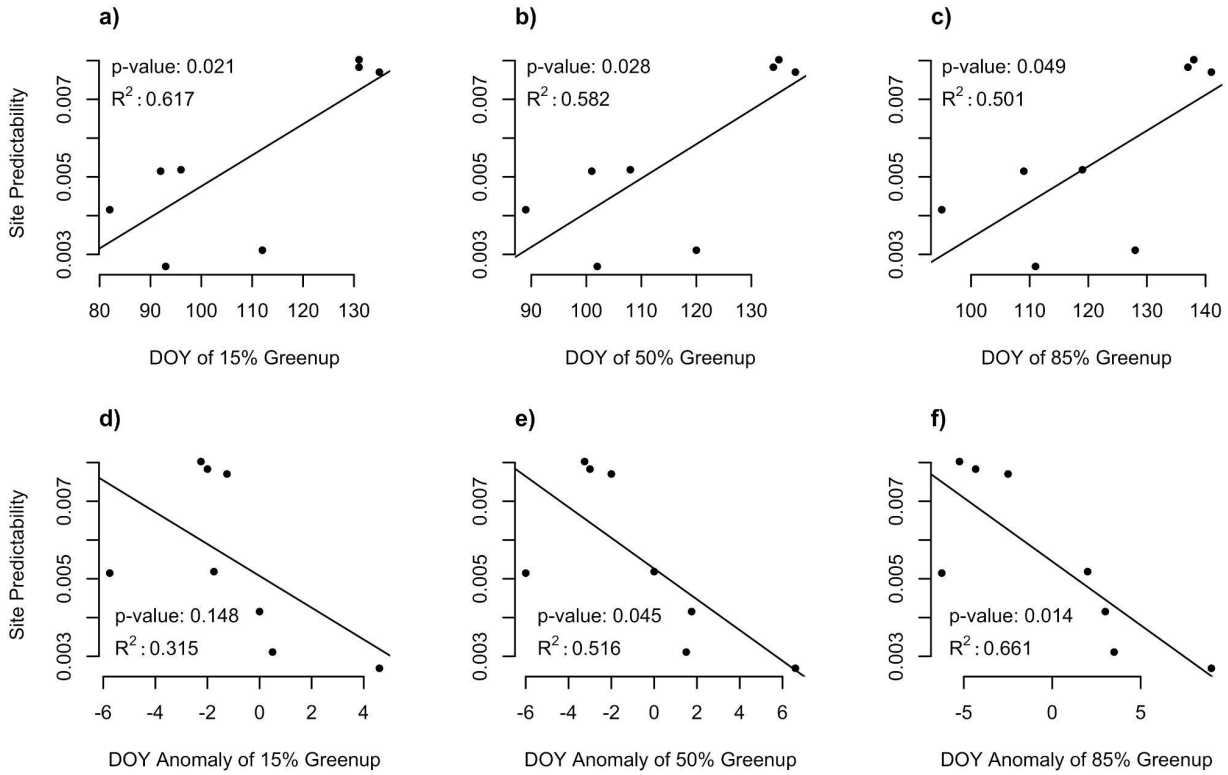

Figure S9. Linear regressions between the site predictability assessed via Continuous Ranked Probability Score and different explanatory variables related to timing with the *p*-value and coefficient of determination (R^2) given in each panel. The explanatory variables tested were the day of year (DOY) of 15% (a), 50% (b), and 85% greenup (c) and the day of year anomaly of the 15% (d), 50% (e), and 85% (f) greenup days. Variation in predictability amongst sites was significantly explained by DOY of 50% and 85% greenup and the anomaly in the timing of 50% and 85% greenup.

Appendix: Model Descriptions

Below are the model descriptions provided by the participants of the Challenge and edited for consistency.

CSP-Gwave

The CSP-Gwave ("greenwave") model represented seasonality using a double logistic function. Within a single year (growing season), the function consisted of two primary components: one curve for greenup, and another for "brown-down". It accommodated trends in greenness between periods of greenup and brown-down by allowing greenness to increase or decrease as vegetation (*e.g.,* leaf physiology) changes over the winter or summer (Elmore et al., 2012). It used a corrected sum of the curves for each year to obtain a single, continuously differentiable curve spanning the time period of interest. Spatial and temporal heterogeneity was accommodated using fixed and random effects on the parameters of the curve. Most, if not all, of the parameters in the model had direct, phenological interpretations – metrics such as the timing of peak growth, peak greenness, and duration of the growing season. The greenwave model was developed originally to handle the sparsity of observations in Landsat time series of vegetation indices by "borrowing strength" from observations in all sites and years. One major limitation of the model in the phenology forecasting challenge – and its application to PhenoCam data, generally – is that covariates were not allowed to vary daily. The team did not expect this model to compete favorably with models that are able to use more granular covariate information.

The greenwave model was implemented using greta in R on a GPU-enabled VM hosted on Azure. Most of the forecasts were made by hand or using simple cron jobs. Training data included all available deciduous forest NEON sites: HARV, BART, SCBI, STEI, UKFS, GRSM, DELA, CLBJ, BLAN, TREE, UNDE, MLBS, ORNL, and LENO. The final model included spatial random effects on greenness during the winter season, and on the 'greendown factor' — the parameter used to control changes in greenness during the growing season. Spatiotemporal random effects were placed on parameters related to the timing of peak greenup, the magnitude

of peak greenness, and the duration of the growing season. Fixed effects on parameters included mean annual precipitation, temperature, a precipitation-by-temperature interaction term, long-run summary stats (quantiles) for Normalized Difference Vegetation Index (NDVI) at each site developed using MODIS, and field latitude. All model refinements involved post-hoc analysis to determine if as-yet unincorporated site level biophysical factors explain variability in random effects on parameters. Additional details on the model, including all model math, can be found here: https://gitlab.com/apis-staging/greenwave (on the "neon" branch).

Team members included Luke Zachmann and Vincent Landau.

CU_Pheno

The CU_Pheno model was a deterministic, discrete time compartment model that simulated the transition of pixels between green (G) and non-green (N) color channels as the forest moves through yearly transitions. The model was initiated at an initial greenness value, G_init, based on the previous three days of data at any point in time. Three epochs with distinct growth rates were considered: winter dormancy, which had no significant growth or decay; spring greenup, modeled by steep linear growth; and summer leaf maturation, modeled by an inverse exponential decay function. The transition from winter to spring greenup was based on the number of growing-degree days (GDD's) accumulated since January 1 of the given year, using NEON's temperature data. The transition from spring greenup to summer leaf maturation occurred when the maximum G_{CC} value was reached, informed by the historically observed data. Five parameters were fit to all existing historical data: initial G_{CC}, spring growth rate, summer decay rate, winter-spring transition (GDD), and spring-summer transition (G_{CC} maximum). The parameter combination that minimized the sum of the squares when comparing model

predictions to historical data for each site was used for forecasting. A scaled k-fold cross validation method was used to predict the model uncertainty. Additional details can be found at https://github.com/lcpowers/PhenoRepo.

Team members included Josh Seabaugh, Casey Middleton, Claire Powers, Brett Melbourne, and Eric Vance.

DALEC-SIP

The DALEC-SIP model was a process-based carbon cycle and ecosystem model. It coupled a leaf-to-canopy radiative transfer model based on the spectral invariant properties (SIP) to the Data Assimilation Linked Carbon (DALEC) model. The DALEC model simulated detailed processes of vegetation phenology, carbon fluxes and carbon pool dynamics. The SIP-based radiative transfer model physically links fundamental leaf traits (*e.g.,* pigment concentrations) and canopy structural parameters (*e.g.,* Leaf area index, clumping index) to the observable land surface reflectance and vegetation indices (*e.g.,* G_{CC}).

In the current phase of forecasting challenge, the DALEC-SIP model was setup and manually calibrated at each site with a variety of data from literature review, including Leaf Mass per Area (LMA), Vcmax, chlorophyll content, brown pigment content, peak growing season LAI, and eddy covariance carbon fluxes. Some important parameters that are not available, such as leaf angle distribution, clumping index and the PhenoCam-target observation geometry, were assumed with typical values by experience. Additional details can be found at https://github.com/hliu666/DALEC_SIP/tree/main.

Team members included Haoran Liu, Min Chen, and Dalei Hao.

EFI_U_P

The EFI_U_P model, as part of an undergraduate course in ecological forecasting at Virginia Tech, was a logistic growth process model (Eq. S1) where greenness (G_{CC}) is represented as a function of the parameters ($\theta_1, \theta_2, \theta_3, \theta_4$) and day of year ($x_t$). G_{CC} was forecasted for a 35-day period at the PhenoCam NEON sites (BART, CLBJ, DELA, GRSM, HARV, SCBI, STEI, UKFS).

$$G_{CC,t} = \theta_1 + \theta_2 * \frac{exp(\theta_3 + \theta_4 * x_t)}{1 + exp(\theta_3 + \theta_4 * x_t)}, \quad \text{Eq. (S1),}$$

Parameters were calibrated using a Bayesian framework, with a Monte Carlo Markov Chain (MCMC) simulation. To parameterize, uninformed priors and initial values were used. θ_3 was not parameterized using MCMC and its value was set at -50. Within the MCMC framework, 10,000 iterations were run with 3 chains for each parameter. A burn-in period was set for 1000 iterations.

To quantify total uncertainty for the forecast, the standard deviation around the calculated mean values was quantified and visualized. Due to the methods used in this forecast, total uncertainty encompasses both process and parameter uncertainties. Additional model details can be found on github: https://github.com/jacob8776/EFI_U_P_challenge**.**

Team members included Jacob Wynne.

Fourier

The Fourier model captured the time-dependence of the greening process using the Fourier regression method and a set of temporal basis functions given below

$$f_n(t) = cos\,(2\pi nt/T), \quad \text{Eq. (S2),}$$

$$g_n(t) = sin\,(2\pi nt/T), \quad \text{Eq. (S3),}$$

where $T = 365$ days, and $n \in \{0, 1, \ldots 50\}$ for f_n and $n \in \{1,2, \ldots 50\}$ for g_n, giving us the final fitting function,

$$f(t) = a_0 + \sum_{n=1}^{50} a_n \cos(2\pi nt/T) + b_n \sin(2\pi nt/T), \qquad \text{Eq. (S4).}$$

The team fit the coefficients $\{a_n, b_n\}$ using ordinary least squares regression method on the sinusoidally transformed variables, $\{cos\ (2\pi nt/T), sin\ (2\pi nt/T)\}$. The advantage of using OLS on basis functions was that it was extremely fast and robust to missing time points. This method may be expanded to include weather features as additional variables to the OLS regression. The Fourier team chose to omit weather as they were aiming to perform long term forecasts (beyond 7–14 days horizon where weather forecasts are generally unavailable).

Uncertainties in the predictions were calculated by fitting the model (Eq. S4) to the squared responses, and estimating the prediction variance as,

$Var[Y] = \widehat{Y^2} - \widehat{(Y)}^2$ where $\widehat{Y^2}$ and Y are estimated from the data.

Additional details can be found at https://github.com/uttambhat/ecology_forecast_challenge_2021.

Team members included Uttam Bhat.

greenbears (greenbears_gams, greenbears_stl, and greenbears_par)

The “greenbears” team explored a variety of simple parametric models for forecasting phenology change at the eight target competition sites. The greenbears team participated in the EFI challenge as part of a graduate seminar at University of California, Berkeley led by Carl Boettiger titled “Reproducible & Collaborative Data Science”.

Over the course of the challenge, the greenbears team submitted forecasts from several different models. They started with a simple parametric approach and tried to fit a smooth cyclic

curve based on day-of-year for each site. They used generalized additive models (GAMs) to estimate an annual smooth cycle in the G_{CC} greenness index (Wood, 2017). The first model they submitted (as *greenbears_gams*) assumed every year should have the same cycle and provided no covariates. They explored forecasting with a "Seasonal and Trend decomposition using Loess" (STL; Hyndman and Athanasopoulos, 2021) model but found that it performed worse than the GAM approach, so they decided to refine their GAM model rather than pursue the STL approach. They performed exploratory analysis on different variables that might be correlated with G_{CC} values, given the historical data that was available. Based on this, they chose to incorporate photosynthetically active radiation (PAR) into our second GAM model (NEON, 2021). Therefore, their model *greenbears_par* extended the GAM design with a linear effect of PAR. They generated a PAR forecast based on its historical cycle using a first-stage cyclical GAM on historical PAR values, an ad hoc approach that they hypothesized might capture a different aspect of the cyclical trend. Throughout all approaches, they treated the eight target sites as independent.

To automate this code, they set up a Github Action and wrote a script that pulled up-to-date phenology and covariate data, estimated models, and submitted those models to the NEON API. They then scheduled this code to run every morning and submit a forecast. An archived snapshot of the repository is available on Zenodo at http://doi.org/10.5281/zenodo.5738488**.**

Team members included Ben Goldstein, McKalee Steen, and Raphaela Floreani Buzbee.

PEG

The PEG model was intended as a simple but informed "proof of concept". Initial versions were historical means with a few different ways of estimating standard deviation.

Around mid-March during the Challenge period, this was changed to use the R 'forecast' package (Hyndman and Khandakar, 2008) with seasonal and exponential smoothing. This model was a Simple Seasonal + Exponential Smoothing Model, with the G_{CC} targets as inputs. Additional details are available through https://github.com/genophenoenvo/neon-efi-challenge/tree/master/simple.

Team members included Debasmita Pal, Jessica Guo, David LeBauer, and Arun Ross.

PEG_RFR (PEG_RFR0, PEG_RFR, PEG_ RFR2)

Three versions using random forest regression were developed as part of this challenge. The models were primarily developed as auto-regressive models based on the past G_{CC} data. For future predictions in case of PEG_RFR2, hourly NOAA weather forecasts were obtained via the EFI website, summarized as the median across 21 model ensembles, and summarized as the mean or sum into daily variables. Code for all PEG models are available at https://github.com/genophenoenvo/neon-efi-challenge/tree/master/ML.

Team members included Debasmita Pal, Jessica Guo, David LeBauer, and Arun Ross.

PEG_RFR0

The PEG_RFR0 model was a multi-output regression model, which predicted 36 days of G_{CC} (to next 35 days) using immediate past G_{CC} (last 5 days) as well as G_{CC} value from the last year (25 days). This meant that to predict t to (t+35) days of G_{CC}, this model used G_{CC} value from (t-1)th to (t-5)th day and G_{CC} value from (t-5)th to (t+19)th day of last year. Each site was modeled individually with Random Forest Regression (RFR) using 3-fold cross-validation after dropping

all the missing values and evaluated based on Root Mean Square Error (RMSE) and R-squared value.

PEG_RFR

The PEG_RFR model was also a multi-output regression model, which predicted 36 days of G_{CC} (t to t+35 days) using immediate past G_{CC} (last 5 days) and G_{CC} value from the last year, meaning (t-5)th to (t+14)th days G_{CC} from last year. The window size of last year was reduced in this case because linear interpolation was used to fill in missing values. Then, a similar process was followed as in PEG_RFR0.

PEG_RFR2

The PEG_RFR2 model was a single output regression model, which used G_{CC} value from last year and current weather data, meaning to predict G_{CC} value of t^{th} day, we use $(t-7)^{th}$ to $(t+7)^{th}$ days' G_{CC} of last year and weather variables (max_temp, min_temp, radiation, precipitation) of t^{th} day. Past weather data was extracted from Daymet and used to train the model. To forecast G_{CC} for future days, the median of NOAA ensemble forecasted weather parameters were used.

PhenoPhriends

The PhenoPhriends model, as part of the ecological forecasting course at Boston University, predicted greenness with a structure based on the logistic growth function and fit using a Bayesian framework. The outcome of the model was a forecast of greenness (G_{CC}) each day for a 35-day window at eight NEON sites (BART, CLBJ, DELA, GRSM, HARV, SCBI, STEI, UKFS). In this model, next day greenness was equal to the current day's greenness plus coefficient term determined by that day's maximum temperature. Inputs to the model included

previous day's G_{CC} value from NEON PhenoCam data and maximum daily temperatures derived from NOAA ensemble weather data provided by the EFI team. Training data for the PhenoPhriends model included historical G_{CC} data and Daymet maximum daily temperatures at each site.

The PhenoPhriends team fit separate process (Eq. S5) and data (Eq. S6) models with uninformative priors to forecast spring G_{CC} and quantify observation, parameter, and process uncertainties.

$$G_{CC,t+1} = N(G_{CC,t}, \tau) + N(maxtemp, \tau), \quad \text{Eq. (S5)},$$

$$Y_t = N(G_{CC,t}, \tau), \quad \text{Eq. (S6)},$$

where τ was given a prior of a Gamma distribution.

They assimilated new G_{CC} observations by fitting an iterative MCMC approach in JAGS using the 'rjags' package in R (Plummer, 2021). Additionally, they applied a Kalman filter for their final forecast to assimilate current-day G_{CC} observations before each new forecast timestep. This allowed them to not only visualize the current predictions for the next 35 days but also how that prediction had changed over the course of time (typically a week). Their code is available at https://github.com/EcoForecast/PhenoPhriends.

Team members included Sam Agate, Devin Hubbard, Mira Kelly-Fair, and Charlotte Malmborg.

TEAM_MODIS

The TEAM_MODIS model predicted greenness based on the spatio-temporal model under the Bayesian framework that the TEAM_MODIS team had previously developed to forecast the peak onset of greenness (POG) of vegetation. The model details can be found in

Neupane et al. (2022). In this model, growing-degree-day (GDD) and land cover types were predictors and the MODIS satellite observed POG was the response variable (Friedl et al., 2012). The daily GDD (base -5°C) was calculated using the daymet daily minimum and maximum temperature values (Thornton et al., 2020). The final model was based on the observations from 2001 to 2016. Using this model, they predicted the mean POG in all NEON sites. Then, they extracted the daily observed temperatures from the North American Land Data Assimilation System (NLDS, Xia et al., 2012) assimilated product for all NEON sites. This data was available at a time lag of about six days. Then, using the NLDS daily temperature data, they calculated the daily accumulated GDD up to the POG as predicted by their model. Then, they pulled the current accumulated GDD values for each site and compared this to the average accumulated GDD from 2001 to 2016 to assess how far off they were from the average GDD value on the same day of year from past years. They used this offset of current versus average past years to shift a natural spline fit to the data by a set number of days. For example, if the current year was lagging 5 days behind the average GDD accumulation for greenup from 2001 to 2016, they shifted our spline values backward by five days. They then used these spline values to inform our predictions. Uncertainty around their predictions was rendered as a constant value by taking the standard deviation of G_{CC} in three distinct phases: (1) initial phase, in which G_{CC} values have not yet risen (between day of year 0 and 60; uncertainty calculated by 2 standard deviations); (2) rising phase, or an active greenup phase (between day of year 61 to 150; uncertainty calculated by 5 standard deviations); and (3) termination phase, or when G_{CC} values no longer rise (after day of year 151; uncertainty calculated by three standard deviations). Additional details can be found at https://github.com/syamghali/Neupane_EtAl_2022_GeoInfo.git

Team members included Leslie Reis, Naresh Neupane, and Vaughn Shirey.

GPEDM

The GPEDM model used an empirical dynamic modeling (EDM) approach. This was chosen because the coupling between plant phenology and environmental variables is highly nonlinear. This approach is based on the mathematical theory of reconstructing system attractors using time-delay embedding (Takens, 1981). It operates with minimal assumptions and reveals complex causal relationships from time series. In practice, it has been used effectively in predicting fish population dynamics, outperforming parametric alternatives. Multivariate EDM can account for the effects of environmental factors. In particular, the GPEDM team used a Bayesian approach to reconstructing system attractors, named Gaussian Process empirical dynamic modeling (GP-EDM) (Munch et al., 2017).

Time series of PhenoCam Green Chromatic Coordinate (G_{CC}) (Seyednasrollah et al., 2019) were pre-processed to fill in gaps with linear interpolation, smoothed with Whittaker smoothing, and transformed linearly within each site to the range of (-0.5, 0.5) to be scale-free. Historical and forecasted meteorological data at focal sites, including daily maximum temperature, daily minimum temperature, and daily precipitation, were retrieved from NEON (DP1.00003.001 and DP1.00006.001) and NOAA. Meteorological data were also smoothed and transformed.

The GPEDM team used time-delayed variables, both G_{CC} itself and meteorological variables, as the predictors (X) of G_{CC} on the day of interest (y) in the GP-EDM. They calculated the mean G_{CC} in windows of 16 days for 8 windows before the day of interest. For each meteorological variable, they calculated their means in windows of eight days for eight windows

before the day of interest. A hierarchical structure was designed to describe how the phenology-environment relationship (*f*) is auto-correlated over space and time.

$$p(y_{s,d,t}|f_{s,d}, X_{s,t}, \epsilon) \sim N(f_{s,d}(X_{s,t}), \epsilon), \quad \text{Eq. (S7)},$$

$$p(f_{s,d}|\phi, \tau, \rho, \omega) \sim GP(g_s, \Sigma_d), \quad \text{Eq. (S8)},$$

$$p(g_s|\phi, \tau, \rho) \sim GP(h, \Sigma_s), \quad \text{Eq.(S9)},$$

$$p(h|\phi, \tau) \sim GP(0, \Sigma), \quad \text{Eq. (S10)},$$

$$\Sigma_d = exp(-\frac{\omega}{2}||d_i - d_j||^2)\Sigma_s, \quad \text{Eq. (S11)},$$

$$\Sigma_s = \rho I(s_i = s_j)\Sigma, \quad \text{Eq. (S12)},$$

$$\Sigma = \tau^2 exp(-\frac{\phi}{2}||X_i - X_j||^2), \quad \text{Eq. (S13)},$$

where φ is the characteristic length-scale for *j*th predictor, governing how much *h* varies in the direction of the *j*th predictor; τ^2 is the variance of *h*; ρ is the spatial correlation, governing how much *g* varies across sites (*s*); ω is the length-scale for day of year (*d*), governing how much *f* varies over time; ε is the Gaussian noise for *y*. Notably, they adopted automatic relevance determination (ARD) by setting the priors of φ to be dependent on the time lag of the *j*th predictor, such that predictors closer to the day of interest are more likely to have larger φ.

For efficient calculation, they used sparse Gaussian Process distributions with a representative set of 500 basis vectors, each drawn from a cluster of historical states from k-means clustering. They used back propagation, a gradient-based method to optimize the model parameters. At each iteration, parameters were chosen to improve the one-step-ahead forecasting accuracy, until convergence or 200 iterations. They implemented a stochastic version of back propagation, optimizing parameters for 100 random observations in every 10 iterations. Every

day, they trained empirical dynamic models with historical phenological and meteorological data and made predictions with optimized parameters and forecasted meteorology in the following 35 days. Additional details can be found at https://github.com/zhulabgroup/phenology-efi/.

Team members included Yiluan Song, Stephen B. Munch, Uttam Bhat, and Kai Zhu.

VT_Ph_GDD

The VT_Ph_GDD model, as part of a graduate course in ecological forecasting at Virginia Tech, was built using a logistic growth process model (Eq. S14) fit in a Bayesian framework to predict spring greenness, where x_t = cumulative growing degree day (GDD) on the day being predicted. GDD was calculated using Eq. (S15) on each day a forecast was made, where $max_{temp}$is the maximum air temperature and $min_{temp}$is the minimum air temperature on a given day, and $T_{base}$was a constant set at 10°C.

$$GCC_t = \Theta_1 + \Theta_2 \frac{exp(\Theta_3 + \Theta_4 x_t)}{1 + exp(\Theta_3 + \Theta_3 x_t)}, \quad \text{Eq. (S14),}$$

$$GDD = \sum max((max_{temp} + min_{temp})/2 - T_{base}), 0), \quad \text{Eq. (S15),}$$

To forecast the covariate, GDD, 35 days into the future the VT_Ph_GDD team used the National Oceanic and Atmospheric Administration (NOAA) Global Ensemble Forecasting System (GEFS) forecast of air temperature in Eq. (S15), taking the average across all 21 ensembles in the NOAA GEFS forecast. Any days in the historical dataset when air temperature was not available through the NEON meteorological data, the VT_Ph_GDD team filled using the NOAA GEFS air temperature forecast for that day.

To quantify model process, observation, and parameter uncertainty, they fit a separate process model (Eq. S14) and data model (Eq. S16), with uninformed priors (Eq. S17) on Θ_1, Θ_2, and Θ_4, with Θ_3 set to -50.

$$y_t = N(GCC_t, \sigma), \quad \text{Eq. (S16)},$$

$$\Theta_x \sim N(0, 10000), \quad \text{Eq. (S17)},$$

Where σ~U(0.0001,100).

They used an iterative batch fitting approach to assimilate new observations of G_{CC}, whereby all data was re-fit using nimble each day a forecast was made, using three chains, 10,000 iterations, and a burn in period of 1,000 iterations. They assessed convergence of chains using the potential scale reduction factor of the Gelman-Rubin statistic ($\hat{R}$), where chains were considered to have converged when $\hat{R}$ was less than 1.2. Additional details can be found at https://github.com/eco4cast-class-VT/phenology_grad.

Team members included Benjamin Miller, Whitney Woelmer, and Garret Dettman.

Supplementary References

Breiman, 2001. Random Forests. Machine Learning, 45(1), 5–32. https://doi.org/10.1023/A:1010933404324

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. Global Change Biology 18, 656–674. https://doi.org/10.1111/j.1365-2486.2011.02521.x

Friedl, M.A., McIver, D.K., Hodges, J.C.F., Zhang, X.Y., Muchoney, D., Strahler, A.H., Woodcock, C.E., Gopal, S., Schneider, A., Cooper, A., Baccini, A., Gao, F., Schaaf, C., 2002. Global Land Cover Mapping from MODIS: Algorithms and Early Results. Remote Sensing of Environment 83, 287–302. https://doi.org/10.1016/S0034-4257(02)00078-0

Hyndman, R.J., Khandakar, Y., 2008. Automatic time series forecasting: the forecast package for R. Journal of Statistical Software 26 (3), 1–22. https://doi.org/10.18637/jss.v027.i03

Hyndman, R.J., Athanasopoulos, G., 2021 Forecasting: principles and practice, 3rd edition, OTexts: Melbourne, Australia. OTexts.com/fpp3. Accessed on February 8, 2021.

Munch, S.B., Poyner, V., Arriaza, J.L., 2017. Circumventing structural uncertainty: A Bayesian perspective on nonlinear forecasting for ecology. Ecological Complexity 32, 134–143. https::/doi.org/10.1016/j.ecocom.2016.08.006

NEON (National Ecological Observatory Network). Photosynthetically active radiation (PAR), RELEASE-2021 (DP1.00024.001). Dataset accessed from https://data.neonscience.org on March 1, 2021. https://doi.org/10.48443/8a01-0677.

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. International Journal of Applied Earth Observation and Geoinformation 108, 102747. https://doi.org/10.1016/j.jag.2022.102747

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, É., 2011. Scikit-learn: Machine Learning in Python. J. Mach. Learn. Res. 12, 2825–2830.

Plummer, M., 2021. rjags: Bayesian Graphical Models using MCMC, available at: https://CRAN.R-project.org/package=rjags

Takens, F., 1981. Detecting strange attractors in turbulence. Lecture Notes in Math 898, 366–381.

Thornton, P.E., Thornton, M.M., Mayer, B.W., Wei, Y., Devarakonda, R., Vose, R.S., Cook, R. B., 2016. Daymet: Daily Surface Weather Data on a 1-km Grid for North America,

Version 3. ORNL DAAC, Oak Ridge, Tennessee, USA. https://doi.org/10.3334/ORNLDAAC/1328

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. Scientific Data 6, 222. https://doi.org/10.1038/s41597-019-0229-9

Wood, S.N., 2017 Generalized Additive Models: An Introduction with R (2nd edition). Chapman and Hall/CRC.

Xia, Y., Mitchell, K. , Ek, M., Sheffield, J., Cosgrove, B., Wood, E., Luo, L., Alonge, C., Wei, H., Meng, J., Livneh, B., Lettenmaier, D., Koren, V., Duan, Q., Mo, K., Fan, Y., Mocko, D., NCEP/EMC, 2012. NLDAS Noah Land Surface Model L4 Hourly 0.125 x 0.125 degree V002, Edited by David Mocko, NASA/GSFC/HSL, Greenbelt, Maryland, USA, Goddard Earth Sciences Data and Information Services Center (GES DISC). Dataset accessed in February 2022. https://doi.org/10.5067/47Z13FNQODKV