


APPLICATION

An integrated phenology modelling framework in R

Koen Hufkens^{1,2,3}  | David Basler³ | Tom Milliman⁴ | Eli K. Melaas⁵ | Andrew D. Richardson^{3,6,7}

¹INRA, UMR ISPA, Villenave d'Ornon, France; ²Department of Applied Ecology and Environmental Biology, Ghent University, Aquitaine, Belgium; ³Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA, USA; ⁴Earth Systems Research Center, University of New Hampshire, Durham, NH, USA; ⁵Department of Earth & Environment, Boston University, Boston, MA, USA; ⁶School of Informatics, Computing and Cyber Systems, Northern Arizona University, Flagstaff, AZ, USA and ⁷Center for Ecosystem Science and Society, Northern Arizona University, Flagstaff, AZ, USA

Correspondence

Koen Hufkens

Email: koen.hufkens@gmail.com

Funding information

Directorate for Biological Sciences, Grant/Award Number: EF-1065029; Federaal Wetenschapsbeleid, Grant/Award Number: BR/175/A3/COBECORE; NSF Macrosystems Biology programme, Grant/Award Number: EF-1065029 and EF-1702697; LabEx COTE MicroMic project; US Department of Energy

Handling Editor: Sarah Goslee

Abstract

1. Phenology is a first-order control on productivity and mediates the biophysical environment by altering albedo, surface roughness length and evapotranspiration. Accurate and transparent modelling of vegetation phenology is therefore key in understanding feedbacks between the biosphere and the climate system.
2. Here, we present the PHENOR R package and modelling framework. The framework leverages measurements of vegetation phenology from four common phenology observation datasets, the PhenoCam network, the USA National Phenology Network (USA-NPN), the Pan European Phenology Project (PEP725), MODIS phenology (MCD12Q2) combined with (global) retrospective and projected climate data.
3. We show an example analysis, using the PHENOR modelling framework, which quickly and easily compares 20 included spring phenology models for three plant functional types. An analysis of model skill using the root mean squared (RMSE) error shows little or no difference regardless of model structure, corroborating previous studies. We argue that addressing this issue will require novel model development combined with easy data assimilation as facilitated by our framework.
4. In conclusion, we hope the PHENOR phenology modelling framework in the R language and environment for statistical computing will facilitate reproducibility and community driven phenology model development, in order to increase their overall predictive power, and leverage an ever growing number of phenology data products.

KEYWORDS

modelling, MODIS land surface phenology, PEP725, PHENOCAM, phenology, R package, USA-NPN

1 | INTRODUCTION

Seasonal leaf development, or vegetation phenology, is strongly linked to seasonal changes in temperature and considered an indicator of climate change. Currently, rising temperatures due to climate change have moved spring forward in time by 2.3 days per decade since the

1970s (Rosenzweig et al., 2007). Vegetation phenology hereby does not only disproportionately influence ecosystem productivity by advancing and delaying the season (Richardson et al., 2010, 2013), it also changes canopy properties such as albedo and atmospheric boundary layer properties (Hollinger et al., 1999; Sakai, Fitzjarrald, & Moore, 1997). As such, models of seasonal leaf development, rigorously

validated against in situ observations, are key to understanding how climate change will affect ecosystem productivity and biophysical vegetation properties.

Luckily, phenology has been recorded by amateurs and professionals, such as national meteorological institutions, supporting contemporary analysis of past or ongoing climate change (Chuine et al., 2004). Recently, individual observations have been formalized into rigorous citizen science efforts through, for example, the USA National Phenology Network (USA-NPN; <https://www.usanpn.org/>; Betancourt et al., 2005) and Project Budburst (<http://budburst.org/>). In addition, automated camera networks (i.e. the PhenoCam network, <https://phenocam.sr.unh.edu/>; (Richardson et al., 2018)) or remote sensing (Zhang et al., 2003) provide a canopy wide continuous way of evaluating the development of vegetation across larger areas in a consistent and continuous fashion (Melaas, Friedl, & Richardson, 2016; White et al., 2009). Numerous studies have demonstrated the value of the PhenoCam-derived Gcc index, a measure of vegetation greenness as percentage green within a digital image, for characterizing the seasonal trajectory of vegetation color and activity (Hufkens et al., 2016; Keenan et al., 2014; Klosterman et al., 2014; Toomey et al., 2015). Similarly, the MODIS MCD12Q2 phenology product has been a proven source of phenological data (Chen, Melaas, Gray, Friedl, & Richardson, 2016).

These observations of vegetation phenology allow us to estimate changes in the timing of vegetation development in response year to year variation in weather, climate change and climate variability (Chuine et al., 2004; Melaas et al., 2016; Vitasse, Porté, Kremer, Michalet, & Delzon, 2009). Most process-based models try to simulate various internal and environmental influences, such as whole plant physiological status (paradormancy), internal factors of developing bud (endodormancy) and external factors driving or suppressing seasonal development (ecodormancy) (Lang, Early, Martin, & Darnell, 1987).

One of the first such ecodormancy models was the growing degree day model as proposed by De Reaumur dating back to 1735. Although vegetation phenology is often driven by temperature multiple additional constraints have been proposed including daylength, chilling degrees, precipitation, relative humidity or vapour pressure deficit (Chuine & Cour, 1999; García-Mozo et al., 2009; Hunter & Lechowicz, 1992; Laube, Sparks, Estrella, & Menzel, 2014; Laube et al., 2013; Xin, Broich, Zhu, & Gong, 2015). Similarly, fall senescence has been modelled, using chilling degree days with additional constraints such as daylength (Archetti, Richardson, O'Keefe, & Delpierre, 2013; Gill et al., 2015; Jeong & Medvigy, 2014). These various models are either used in isolation to address particular physiological questions or included in land surface models to scale phenological processes (Richardson et al., 2011). Model development, in isolation or coupled to larger land surface models, often integrate multiple environmental drivers, which increases model complexity (Chen et al., 2016; Jeong & Medvigy, 2014). Yet, models which include more complex concepts, based upon growing degree days, do not necessarily perform better than a simple regression-based approach. As such, model structures still explain a limited amount of the year-to-year variability, and fail to generalize well (Basler, 2016; Clark, Salk, Melillo, & Mohan, 2014; Fisher,

Richardson, & Mustard, 2007; Linkosalo, Häkkinen, & Hänninen, 2006; Schaber & Badeck, 2003). For example, model studies have shown that biologically "incorrect" models can be parameterized to provide good predictions but lacking any biological representation (Hunter & Lechowicz, 1992). A study by Migliavacca et al. (2012) has shown that between-model differences by the end of the century are almost as large as differences between-climate scenario values. As a consequence, different model assumptions will behave disproportionately different under future scenarios affecting their potential impacts and uncertainties (Migliavacca et al., 2012).

With vegetation phenology as a first-order control on ecosystem productivity, accurate and transparent model predictions of vegetation phenology in a changing climate are key. In order to facilitate easy model comparison and future development of new models, we developed the PHENOR model framework for the R language and environment for statistical computing (R Core Team 2016). The PHENOR R package assimilates four important phenological records across a variety of ecosystem, plant functional types and scales. The assimilated datasets provide extensive coverage in the US and Europe and results can be easily scaled globally using various gridded data products made accessible through the software. Here, we provide a worked example for the PHENOR R package using the recent standardized PhenoCam dataset (Richardson et al. 2018; <http://phenocam.us>) to demonstrate the ease with which a suite of phenological models (Table 1) can be evaluated and scaled up from sites to regions and biomes, and extrapolated in both forecast and hindcast modes.

2 | MATERIALS AND METHODS

2.1 | The PHENOR R package

The PHENOR R package assimilates four important phenological records of either observational, near-surface and satellite remote sensing-based records across a variety of ecosystem and plant functional types. The PHENOR R package combines data from near-surface remote sensing through the PhenoCam network using PHENOCAMR and DAYMETR R packages into a phenology modelling framework, which covers data preparation, model optimization and model visualization and consists of a number of key functions. In addition, data from the USA National Phenology Network (USA-NPN), the Pan European Phenology Project (PEP725) and the MODIS land surface phenology product (MCD12Q2) can be ingested. In the interest of brevity both PHENOCAMR, DAYMETR R packages and the PhenoCam source data are described in Appendix S1.

The `format_phenocam()` PHENOR function combines phenophases (also called transition dates) generated by the PHENOCAMR R package, with the climate data downloaded, using the DAYMETR R package. The function requires the location (path) of the generated phenophase output files, together with parameters specifying the phenophase (direction = rising; with rising for spring or falling for autumn) and the threshold value used (threshold = 25), the Gcc percentile to use (gcc_value = 90, Sonnentag et al. 2012) and the offset as a day-of-year value. The offset is the day-of-year in the previous year on which to start reporting climate data, running until this day in the subsequent year. The function returns

model calibration/validation and driver data a nested list of data frames, used in subsequent model optimization (df, see description of **optimize_parameters()** and **model_calibration()** below).

```
df = format_phenocam(path = "/path/to/phenocamr/phenophases/",
  direction = "rising",
  gcc_value = "gcc_90",
  threshold = 25,
  offset = 264)
```

Similarly, the **format_peg725()** function uses PEP725 observational data together with European E-OBS climate data (Haylock et al., 2008) to compile a consistent calibration/validation dataset for European observational records (e.g. Basler, 2016). Data can be downloaded using the **download_peg725()** function. We provide similar functionality for the USA-NPN data. Data can be downloaded through the USA-NPN application programming interface, using **download_npn()** and correctly formatted with **format_npn()**. Furthermore, the **format_modis()** function correctly formats a directory of MODIS MCD12Q2 land surface phenology data (i.e. phenophases, Zhang et al., 2003) as downloaded with the MODISTools R package (Tuck et al., 2014).

Spatially scaling of model results is facilitated through a number of functions. The **format_daymet()** function uses gridded pre-processed Daymet tiles to generate spatially explicit driver data (**download_daymet_tiles()** and **daymet_tmean()** functions of DAYMETR, Appendix S1). The **format_eobs()** function provides the same functionality for the E-OBS climate data. Yet another source of hindcast data is compiled using the **format_berkeley_earth()** function, which allows the user to subset 1×1 degree global historical climate data for any year since 1850 through the Berkeley Earth project (Rohde et al., 2012). Similarly, the **format_cmip5()** function formats 1/4th degree NASA Earth Exchange (NEX) global gridded Coupled Model Intercomparison Project (CMIP5) data of historical reanalysis and representative concentration pathway (RCP 4.5 and RCP8.5) projections. Unlike **format_phenocam()** or **format_modis()** no calibration/validation data is included in the gridded spatial data.

The resulting dataset of all formatting functions is a nested list with the following layout:

```
phenor_data_structure
├── doy (vector)
├── site (vector, NULL for spatial data)
├── location (matrix, latitude and longitude by row)
├── ltm (vector, value: degrees C)
├── Ti (matrix, columns: years, rows: doy, value: degrees C)
├── Tmaxi (matrix, columns: years, rows: doy, value: degrees C)
├── Tmin (matrix, columns: years, rows: doy, value: degrees C)
├── Li (matrix, columns: years, rows: doy, value: hours/day)
├── VPDi (matrix, columns: years, rows: doy, value: Pa)
├── Pi (matrix, columns: years, rows: doy, value: mm/day)
├── transition_dates (vector, NULL for spatial data)
├── georeferencing (NULL for PhenoCam, MODIS, USA-NPN or PEP725 data)
├── size (size of the spatial data)
├── extent (extent of the spatial data)
└── projection (projection of the spatial data)
```

Within the PHENOR data structure, the top level is a particular site. For each site, critical parameters such as the day-of-year range (doy, as specified by the offset), the geographic location (or georeferencing) and matrices holding, minimum temperature (T_{\min}), maximum temperature (T_{\max}) and mean daily temperature (T_i), precipitation (P_i), vapour pressure deficit (VPD_i), daylength in hours (L_i) and calibration/validation transition date (transition_dates) data are provided. Matrices are organized with columns representing a given year, and rows representing a given day-of-year.

Other data are represented as vectors matching the number of columns present in the climate data matrices. Where necessary, data are truncated to match the available climate data. When certain data sources are missing the content of a field is set to NULL.

The **optimize_parameters()** function allows for the easy optimization of model parameters. This function uses two common optimizers, GenSA (Xiang, Gubian, Suomela, & Hoeng, 2013) and rgenoud (Mebane & Sekhon, 2011). The GenSA algorithm combines both the Boltzmann machine and faster Cauchy machine simulated annealing approaches for fast optimizations (Tsallis & Stariolo, 1996), while the genoud routine combines an evolutionary algorithm with a derivative-based (quasi-Newton) method to solve difficult optimization problems (Mebane & Sekhon, 2011). To optimize a calibration/validation dataset (df), one specifies a particular model (e.g. the Thermal Time model, TT), a defined optimizer (e.g. GenSA), an objective function such the root mean squared error (RMSE), upper-lower parameter limits and parameter starting values, when required. Additional control parameters, such as the maximum number or iterations (e.g. max.call), can be provided using a list of options to the control parameter. An example function call to optimize a the TT model is provided below.

```
optimal_par = optimize_parameters(par = NULL,
  data = df,
  cost = rmse,
  model = "TT",
  method = "GenSA",
  lower = c(1;5;0),
  upper = c(365;10;2000),
  control = list(max.call = 40000))
```

Final predicted values for the optimized parameters can be retrieved by running the **estimate_phenology()** function with the optimized parameters.

```
results = estimate_phenology(par = optimal_par$par,
  data = df,
  model = "TT")
```

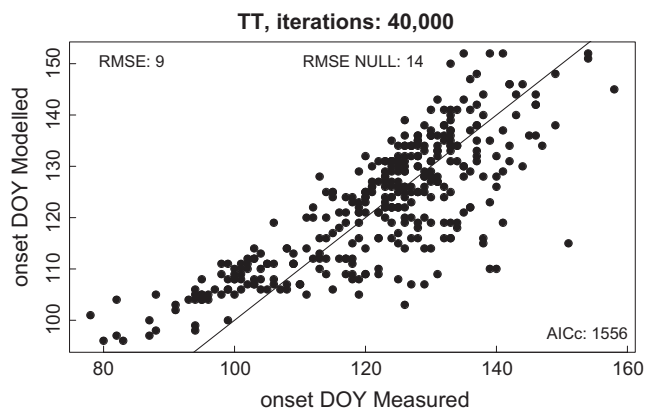


FIGURE 1 Output of the **model_calibration()** function in the PHENOR R package which produces a scatter plot of measured and modelled budburst dates. In this case an optimization was run for the deciduous broadleaf data on a Thermal Time model, using 40,000 iterations of the generalized simulated annealing routine, the package default optimizer. Model fit statistics such as the RMSE and AIC are provided in the top right and bottom left corner of the graph and on the command line output. RMSE NULL is the RMSE assuming the mean of the measured values as the optimal model output. RMSE, root mean squared error; AIC, akaike information criterion

The output will automatically be formatted as a map of phenology dates or a vector, depending on the input data class. However, running models across all grid cells of spatial data would provide a naively broad

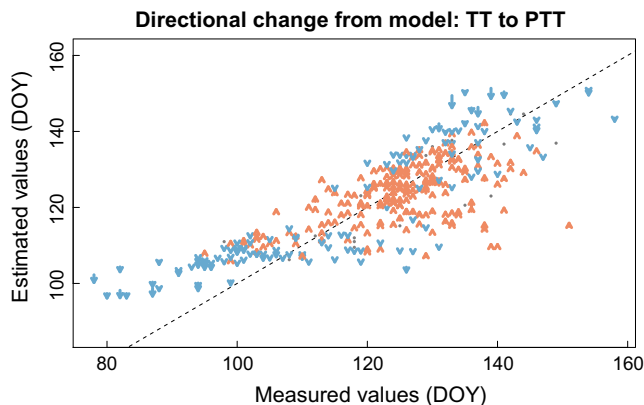


FIGURE 2 Output of the `arrow_plot()` function in the `PHENOR` R package which produces a scatter plot of measured and modelled budburst dates showing directional changes between model runs of the thermal time and photo thermal time models. The direction and magnitude of the change is indicated by arrow's colour and the length, respectively. When changes delay phenological development arrows are coloured orange, when phenological development is advanced arrows are coloured blue. Values unchanged between models structures are indicated by small black dots

TABLE 1 Adapted from Basler (2016) Table 1)

Model Name	Drivers	# Parameters	References/comments
NULL		1	Mean date of leaf unfolding
LIN	F	2	Linear regression against spring temperatures
Ecodormancy release only			
Thermal Time ^a (TT, TTs)	F	3 (4)	(Cannell & Smith, 1983; Chuine, Cour, & Rousseau, 1999; De Reaumur, 1735; Hänninen, 1990; Hunter & Lechowicz, 1992; Kramer, 1994; Leinonen, Repo, & Hänninen, 1997; Wang, 1960)
Chilling Degree Day (CDD)	C	3	(Jeong & Medvigy, 2014)
Photothermal-time ^a (PTT, PTTs)	PF	3 (4)	(Črepinšek, Kajfež-Bogataj, & Bergant, 2006; Masle, Doussinault, Farquhar, & Sun, 1989)
M1 ^a (M1s)	PF	4 (5)	(Blümel & Chmielewski, 2012)
Endo- and ecodormancy releases			
Alternating (AT)	CF	5	(Cannell & Smith, 1983; Murray, Cannell, & Smith, 1989)
Sequential ^b (SQ, SQb)	CF	8	(Hänninen, 1990; Kramer, 1994)
Parallel (PA, PAb)	CF	9	(Hänninen, 1990; Kramer, 1994; Landsberg, 1974)
Unified (UN)	CF	9	(Chuine, 2000)
Sequential M1 ^b (SM1, SM1b)	CPF	9	Combination with the M1 model
Parallel M1 ^b (PM1, PM1b)	CPF	10	Combination with the M1 model
Unified M1 (UM1)	CPF	10	Combination with the M1 model
Growing Season Index ^c (SGSI, AGSI)	FPV	7	(Xin et al., 2015)
Grassland Pollen model (GRP)	FR	5	(García-Mozo et al., 2009)

Overview of the phenological models for leaf unfolding and leaf senescence, and pollen release included in this study. The models are grouped by implemented processes and drivers: chilling temperatures (C), forcing temperatures (F), photoperiod (P), precipitation (R), and vapour pressure deficit (V). Function names in round brackets while full model structures are listed in Appendix S2 of Table 1.

^aAlso calibrated using a sigmoid temperature response (Hänninen, 1990; Kramer, 1994), adding one parameter.

^bAlso calibrated using a bell-shaped chilling response (Chuine, 2000).

^cCalibrated using a cumulative response rather than the rolling mean.

representation of land surface phenology. For example, only a small subset of the US is dominated by any particular plant functional type (PFT), such as deciduous broadleaf forests. In order to better differentiate between different dominant PFT, we include a function `land_cover_density()` which calculates the percentage coverage of a particular MODIS MCD12Q1 IGBP land cover class (Friedl et al., 2010) within a given raster cell for a given location (i.e. CMIP5 data, see Figure 5a,b).

A wrapper function, `model_calibration()`, is provided for both the `optimize_parameters()` and `estimate_phenology()` functions which integrates the previously described steps providing both summary statistics (RMSE and AICc) and a plot (Figure 1) of the model fit. Likewise, the `model_comparison()` function serves as a wrapper for multi-model parameter optimization runs. For a visual comparison limited to two model runs we provide an arrow plot function, `arrow_plot()`, displaying directional changes in the modelled values between the model outputs (Figure 2).

2.2 | A worked example: a quick model comparison

As a worked example we partially recreate the spring phenology model comparison by Basler (2016), using PhenoCam data. However, we note that a similar exercise could be executed with any of the other phenology data sources available through `PHENOR`. The model structures included in this worked example can be described by the

three broad categories: (1) as simple linear regression to spring temperature, (2) models explaining ecodormancy release only, (3) models explaining the release of endo- and ecodormancy. A reference NULL model assumes a fixed mean date of leaf unfolding.

A total of 22 phenology models are included in the package (Table 1). These include 20 spring phenology models including precipitation driven models, one fall senescence chilling degree day model and one grassland pollen release model. In our worked example of the PHENOR R package, we will focus only on the 20 spring phenology models. A full list of the model structures and parameter ranges for the models are provided in Table 2 of Appendix S2 and included in the PHENOR library https://github.com/khufkens/phenor/blob/master/inst/extdata/parameter_ranges.csv.

For this study, we combined spring phenology dates based on PhenoCam 3-day summary data from the standardized PhenoCam

Dataset (Richardson et al. 2018) with Daymet data (Thornton et al., 2017) for three common PFTs, deciduous broadleaf forests, evergreen needleleaf forest and grasslands. A total of 102 sites and 508 site years were included in our calibration/validation dataset, of which 63 were deciduous broadleaf forest sites (358 site years), 18 were evergreen needleleaf forest sites (63 site years) and 21 were grasslands (88 site years). Deciduous broadleaf sites which are moisture limited, and all sites outside Daymet coverage, were excluded from our analysis. The final selected sites span a large geographic area ranging from New Mexico to Southern Alaska, and Maine to California (Figure 3a).

We acknowledge that phenological development as measured, using PhenoCam data represent different physiological processes for different PFTs. For example, the phenology of deciduous forests or grasslands is closely linked to the development of new leaf tissues (Hufkens et al., 2016; Keenan et al., 2014) where evergreen forest

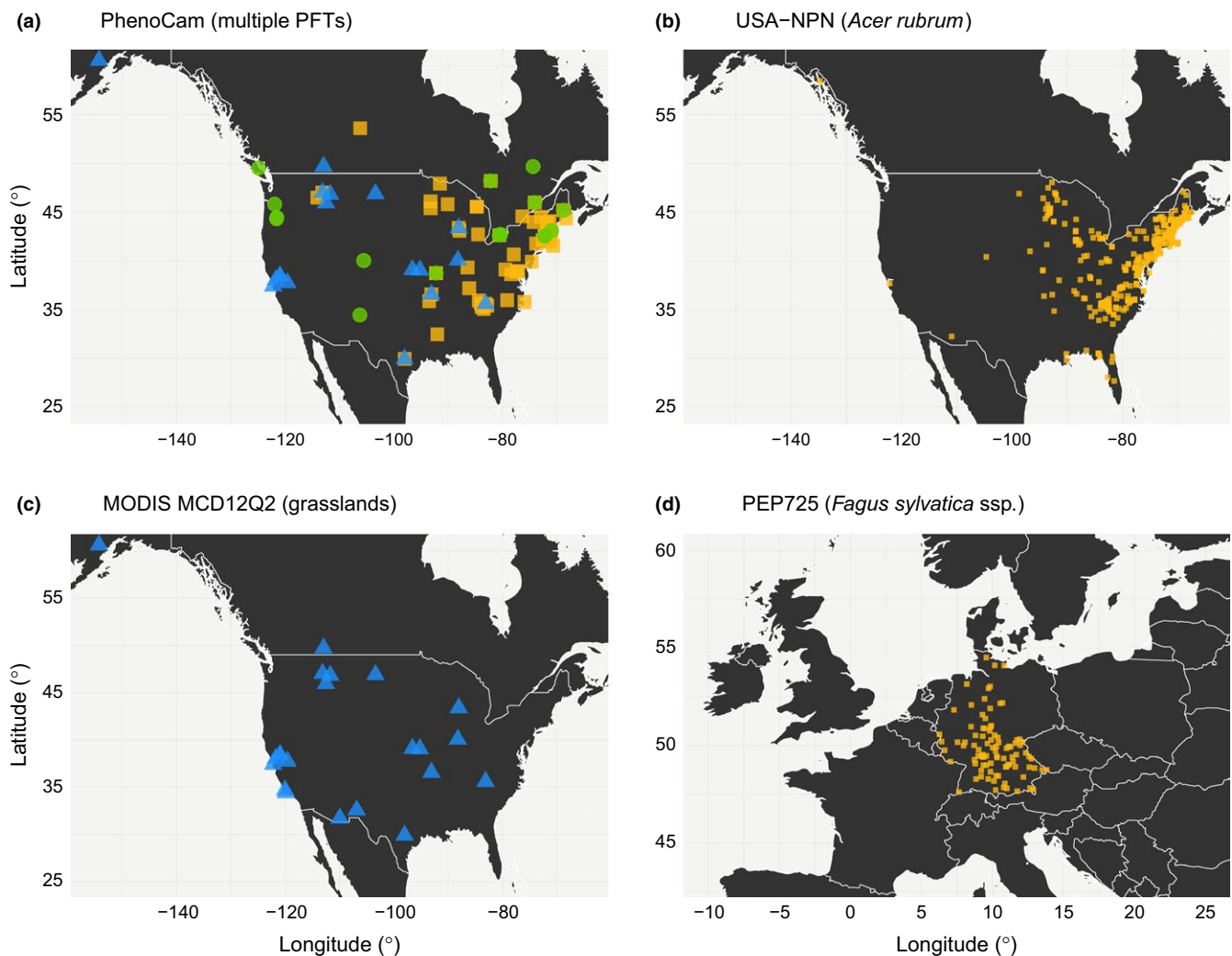


FIGURE 3 Various data products which the PHENOR R package can query for model calibration/validation, and their respective demonstration dataset locations. (see `demo_data.r` to compile this dataset) (a) All locations of the selected 104 PhenoCam Dataset V1.0 sites as used in the worked example. (b) Example locations of all *Acer rubrum* locations for which "Breaking Leaf Buds" observations were made within the USA National Phenology Network. (c) Example locations of MODIS MCD12Q2 grassland phenology colocated with the PhenoCam grassland sites. (d) All *Fagus Sylvestica* ssp. observations of BBCH code 11 or "first true leaf" for sites where at least 60 observation exist. The three plant functional types displayed, deciduous broadleaf forests, evergreen needleleaf forests and grasslands, are marked with open yellow squares, green circles and blue triangles respectively

phenology is determined by dehardening of existing needles at the end of the winter season. Thus, optimized model parameters and their interpretation are specific to each PFT.

For all PFTs, model optimization was executed using the default generalized simulated annealing (GenSA) package and algorithm minimizing the RMSE between the greenness rising PhenoCam phenophase estimations and model predictions (see Appendix S1). The optimizer was run for 40,000 iterations with a starting temperature of 10,000. To determine the influence of locations at the margin of the forest biome on model optimizations, a subset of sites centrally located within the deciduous forest biome was created (Melaas et al., 2016, Appendix S2 Table 2). This subset was optimized separately and compared to results for the complete deciduous broadleaf dataset. We assess proper convergence of the optimized parameters by initializing the optimizer, using 12 random sets of parameters. We report mean and standard deviations of the RMSE between observations and predictions on the optimized parameter values for all datasets. We compare model performance with a log transformed ANOVA, combined with a post hoc Tukey HSD test. Model errors are evaluated for normality, using a Shapiro–Wilk test.

For illustrative purposes, we produce overview maps (Figure 5) of spatial patterns both in hindcast and forecast mode. In hindcast mode, we use 1×1 km Daymet gridded data across New England, while we present the difference in predicted spring phenology (Δ DOY) between

years 2099 and 2011 for forecast CMIP5 IPSL-CM5A-MR (Mid-Resolution Institut Pierre Simon Laplace Climate Model 5) model runs, across the contiguous US. The Thermal Time (TT) and Accumulated Growing Season Index models were optimized for deciduous broadleaf and grassland PhenoCam data, respectively. For forecast data only pixels dominated by their particular PFT (>50% coverage) are displayed, limiting a naively broad interpretation of the results.

Our comparison of 20 spring phenology models across three PFTs showed that most models were significantly different from the NULL model, with the exception of the SGSI model in the evergreen PFT (post hoc Tukey HSD test, $p < .001$, Figure 4). The model performance of the centrally located deciduous broadleaf sites was marginally greater (RMSE: $c. 7.6 \pm 0.7$ days) compared to the complete deciduous broadleaf dataset (RMSE: $c. 7.9 \pm 1.2$ days). This difference between the full deciduous broadleaf forest dataset and a subset of more centrally located sites within the biome suggests an influence of geographic location on model error.

The influence of different model structures on individual values was visualized using the arrow plot (Figure 3) between two model runs. When visually comparing the Thermal Time (TT) and the Photothermal Time (PTT) models small changes are noted (Figure 3). Both models accumulate growing degree days where the PTT model adds a photoperiod component to the original TT model. In the PTT model, leaf unfolding is therefore in part dependent on a daylength requirement in

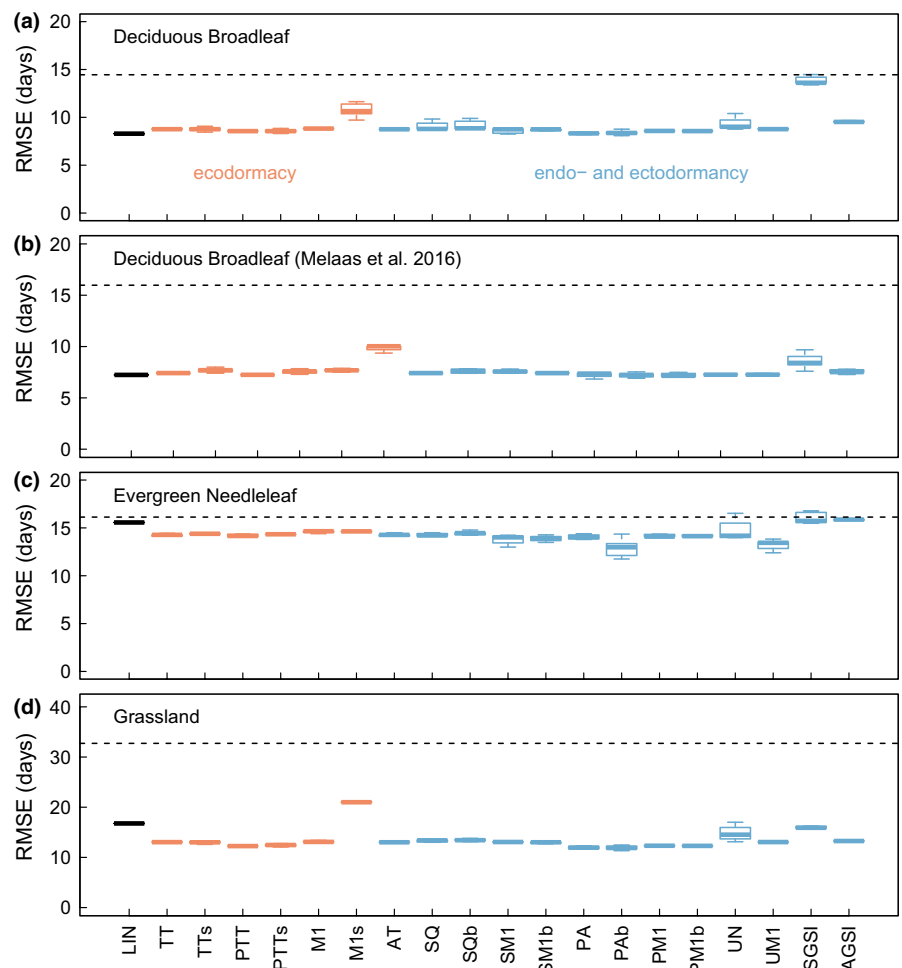


FIGURE 4 Model comparison for 20 spring phenology models and three main plant functional types in the PhenoCam Dataset V1.0 (deciduous broadleaf, evergreen needleleaf and grassland). An additional subset of the deciduous broadleaf dataset is created, matching the PhenoCam locations used in Melaas et al., (2016). All panels show boxplots of the root mean squared error of measured versus predicted budburst dates (in days) for the 20 models used in this study. Models accounting for ecodormancy influences are marked in orange. Models accounting for endo- and ectodormancy are marked in blue. The linear model is marked in black, while the NULL model is represented by the dashed horizontal line

addition to thermal forcing. In our example, the addition of a daylength requirement shifts model results for early and late developing plants toward earlier leaf out dates, while at the same time, shifting mid-season developing plants towards later leaf out dates. Despite these changes, the overall model accuracy remains the same. A description of all statistical results is provided in Appendix S1, Section 5.

3 | DISCUSSION

Here, we have demonstrated how the `PHENOR` R package and its included “model zoo”, together with consistent estimates of vegetation phenology through PhenoCam network (e.g. `PHENOCAMR`, Appendix S1) or other phenology data sources can be leveraged for a fast and

transparent model comparisons. More so, easy access to various gridded data sources allows for quick spatial scaling of optimized models in both hindcast and forecast mode (Figure 5). For example, the code required to partially reproduce a study by Basler (2016) relied on a mere 15 R commands (see `run_model_comparison.r` in the `PHENOR` manuscript github repository), while the models used are easily readable and well documented. Furthermore, adding model structures is easy compared to other frameworks which rely on either low level languages, are closed source or do not work cross-platform (Brown et al., 2014; Chuine, Garcia Cortazar Aza, Kramer, & Hänninen, 2013; Hänninen & Kramer, 2007). More so, to execute our complete case study, reasonable processing times were recorded (c.48 CPU hours on a recent desktop workstation) although relying on a slower scripting language. Computational loads for data generation and processing at a

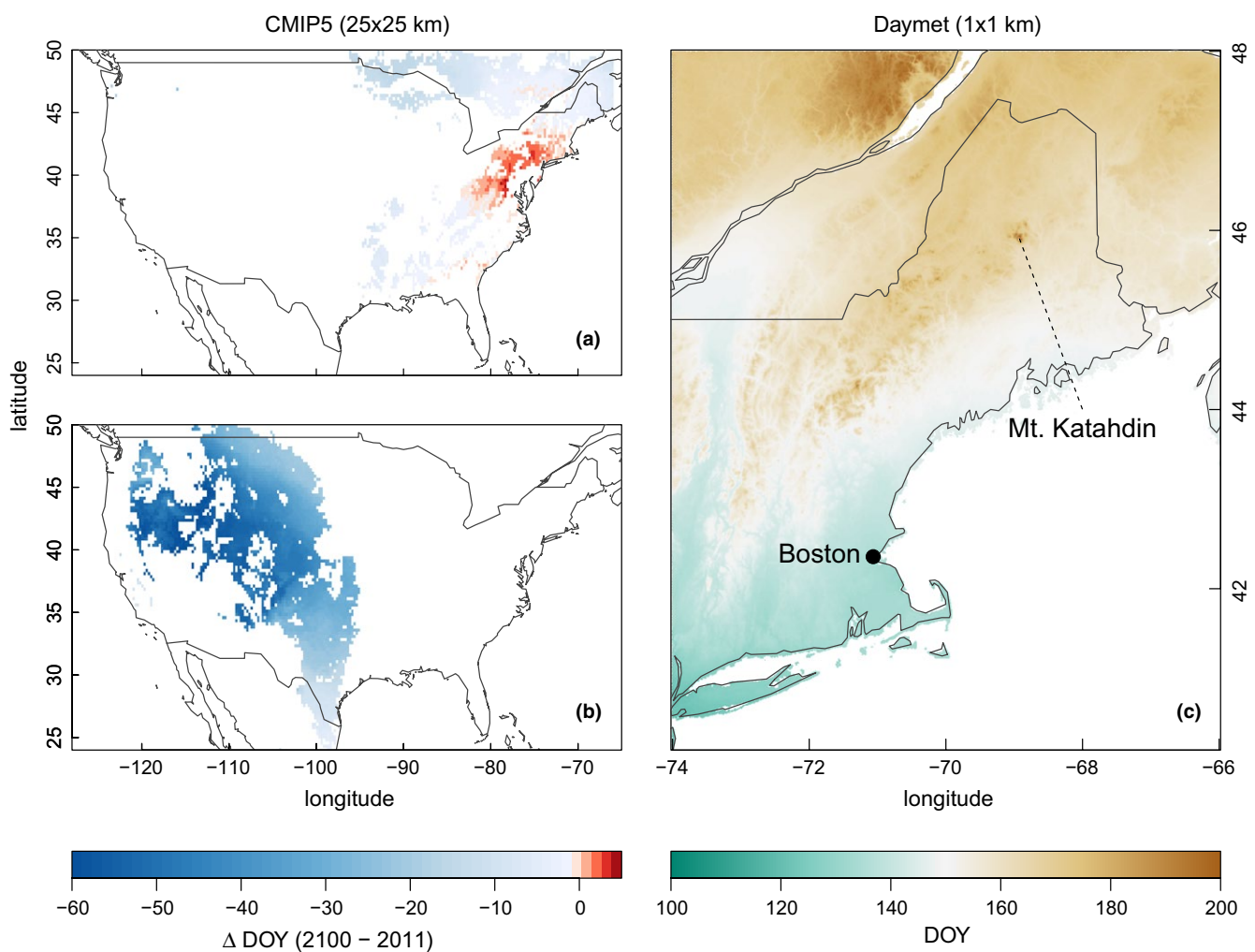


FIGURE 5 Overview map comparing various spatial outputs of the TT and AGSI model optimized to the deciduous broadleaf and grassland PhenoCam data respectively. (a) `PHENOR` model output of the difference in estimates of spring phenology between the year 2100 and 2111 for 1/4th degree NASA Earth Exchange (NEX) global gridded CMIP5 Mid-Resolution Institut Pierre Simon Laplace Climate Model 5 (IPSL-CM5A-MR) model runs, using the TT model parameterized on deciduous forest PhenoCam sites. Only pixels with more than 50% deciduous broadleaf or mixed forest cover per 1/4th degree pixel, using MODIS MCD12Q1 land cover data, are shown; (b) `PHENOR` model output of the difference in estimates of spring phenology between the year 2099 and 2111 for NEX CMIP5 IPSL-CM5A-MR model runs using the AGSI model parameterized on grassland PhenoCam sites. Only pixels with more than 50% grassland coverage per 1/4th degree pixel, using MODIS MCD12Q1 land cover data, are shown; (c) `PHENOR` model output for 11 Daymet gridded datasets (tiles) for the year 2111. TT, Thermal Time; AGSI, Accumulated Growing Season Index, CMIP5, Coupled Model Intercomparison Project 5

global scale remained marginal. The case study demonstrates the ease with which we executed our model comparisons in *PHENOR*, corroborating previous studies and once more highlighting the limitations of current model structures in explaining year-to-year variability (Basler, 2016; Clark et al., 2014; Fisher et al., 2007; Linkosalo et al., 2006). This result therefore underscores the need for tools such as *PHENOR* to facilitate easy and transparent model development and comparison. Furthermore, new visualization methods such as the arrow plot might help in this process. The arrow plot (Figure 2) suggests that the assessment of model skill through summarizing values such as the RMSE seem suboptimal, hiding structural differences in model performance. The non-normal distribution of model errors within all models ($p < .001$) suggests as much.

4 | KNOWN LIMITATIONS

We acknowledge that previous efforts have been made to provide phenology model frameworks (Brown et al., 2014; Chuine et al., 2013; Hänninen & Kramer, 2007). Yet, their use and interoperability and scalability is limited due to platform restrictions or their closed source nature. We recognize that the models as currently presented are by no means an exhaustive list of all model structures found in literature. However, the *PHENOR* R package is open source and adding model structures is easy compared to other low level languages (e.g. C/Fortran) and is actively encouraged. We are aware that the *PHENOR* R package does not include all possible phenological climatological drivers or phenological calibration/validation data either, although we stress that access to four larger freely available phenological data sources is provided by the *PHENOR* R package. Similarly, other sources of climate data, such as the ERA-interim reanalysis data (Dee et al., 2011), could be integrated as long as the described data structure is followed.

5 | CONCLUSION

Accurately representing vegetation phenology, a first-order control on ecosystem productivity, under future conditions is key in understanding feedback between the climate and the biosphere. Here, we demonstrated the advantages of the *PHENOR* R package and modelling framework, through a worked example, by quickly and easily comparing 20 spring phenology models and their model skill for three plant functional types. Our results corroborate previous analysis, showing little or no difference in predictive power between models, which suggest convenient tools for further analysis or novel model development are needed to capture current and future phenological changes as well as their underlying physiological processes. We hope the *PHENOR* phenology modelling framework in R will allow for a better integration of observational and experimental data providing opportunities to better understand the environmental factors driving seasonality, and past and future responses of vegetation to climate change and variability.

ACKNOWLEDGEMENTS

The Richardson Lab acknowledges support from the NSF Macro-systems Biology programme (awards EF-1065029 and EF-1702697). D.B. acknowledges the Harvard Forest Bullard Fellowship programme. K.H. acknowledges support from the LabEx COTE MicroMic project and BELSPO Brain programme (project BR/175/A3/COBECORE). We thank ORNL, the Daymet team and Michele M. Thornton for the continued support in developing *DAYMETR*. We thank the World Climate Research Program's Working Group on Coupled Modelling, which is responsible for CMIP, and the climate modelling groups for producing and making available their model output. We are grateful to the US Department of Energy's Program for Climate Model Diagnosis and Inter-comparison provides coordinating support and led development of software infrastructure in partnership with the Global Organization for Earth System Science Portals. Climate scenarios used were from the NEX-GDDP dataset, prepared by the Climate Analytics Group and NASA Ames Research Center using the NASA Earth Exchange, and distributed by the NASA Center for Climate Simulation (NCCS). We thank the NASA Earth Exchange project for making these data available. Finally, we thank our many collaborators, including site PIs and technicians, for their efforts in support of PhenoCam.

AUTHOR'S CONTRIBUTIONS

K.H. conceived and designed all three packages with contributions of D.B.; T.M. developed the application programming interface queried by *phenocamr*; K.H. analysed the data and interpreted the results; K.H., D.B., E.M., T.M. and A.D.R. interpreted the results. K.H. drafted the manuscript. All authors commented on and approved the final manuscript.

CODE AND DATA AVAILABILITY

All three packages, *PHENOR* (<https://doi.org/10.5281/zenodo.1145369>), *PHENOCAMR* (<https://doi.org/10.5281/zenodo.1136008>), *DAYMETR* (<https://doi.org/10.5281/zenodo.1136231>), can be found on the author's personal github (<http://github.com/khufkens>) and are easily installed and loaded in the R statistical software using the following commands.

```
if(!require(devtools)){install.package(devtools)}
devtools::install_github("khufkens/daymetr")
devtools::install_github("khufkens/phenocamr")
devtools::install_github("khufkens/phenor")
```

The data used in the worked examples are freely available as a curated dataset (Richardson et al. 2018). Manuscript data and figures can be generated using the R scripts listed in the manuscript repository (https://github.com/khufkens/phenor_manuscript/, <https://doi.org/10.5281/zenodo.1136233>).

ORCID

Koen Hufkens  <http://orcid.org/0000-0002-5070-8109>

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>
- Basler, D. (2016). Evaluating phenological models for the prediction of leaf-out dates in six temperate tree species across central Europe. *Agricultural and Forest Meteorology*, 217, 10–21. <https://doi.org/10.1016/j.agrformet.2015.11.007>
- Betancourt, J.L., Schwarz, M., Breshears, D.D., Cayan, D.R., Dettinger, M., Inouye, D.W., ... Reed, B.C. (2005). Implementing a U.S. National Phenology Network. *Eos, Transactions American Geophysical Union*, 86, 538.
- Blümel, K., & Chmielewski, F.-M. (2012). Shortcomings of classical phenological forcing models and a way to overcome them. *Agricultural and Forest Meteorology*, 164, 10–19. <https://doi.org/10.1016/j.agrformet.2012.05.001>
- Brown, H. E., Huth, N. I., Holzworth, D. P., Teixeira, E. I., Zyskowski, R. F., Hargreaves, J. N. G., & Moot, D. J. (2014). Plant modelling framework: Software for building and running crop models on the APSIM platform. *Environmental Modelling and Software*, 62, 385–398. <https://doi.org/10.1016/j.envsoft.2014.09.005>
- Cannell, M. G. G. R., & Smith, R. I. I. (1983). Thermal time, chill days and prediction of budburst in *Picea sitchensis*. *Journal of applied Ecology*, 20, 951–963. <https://doi.org/10.2307/2403139>
- 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. *Global Change Biology*, 22, 3675–3688. <https://doi.org/10.1111/gcb.13326>
- Chuine, I. (2000). A unified model for budburst of trees. *Journal of theoretical biology*, 207, 337–347. <https://doi.org/10.1006/jtbi.2000.2178>
- Chuine, I., & Cour, P. (1999). Climatic determinants of budburst seasonality in four temperate-zone tree species. *New Phytologist*, 143, 339–349. <https://doi.org/10.1046/j.1469-8137.1999.00445.x>
- Chuine, I., Cour, P., & Rousseau, D. D. (1999). Selecting models to predict the timing of flowering of temperate trees: Implications for tree phenology modelling. *Plant, Cell and Environment*, 22, 1–13. <https://doi.org/10.1046/j.1365-3040.1999.00395.x>
- Chuine, I., deGarcia Cortazar Atauri, I., Kramer, K., & Hänninen, H. (2013). Plant development models. In M. D. Schwarz (Ed.), *Phenology: An integrative environmental science* (pp. 275–293). Dordrecht, Netherlands: Springer. <https://doi.org/10.1007/978-94-007-6925-0>
- Chuine, I., Yiou, P., Viovy, N., Seguin, B., Daux, V., & Ladurie, E. L. R. (2004). Historical phenology: Grape ripening as a past climate indicator. *Nature*, 432, 289–290. <https://doi.org/10.1038/432289a>
- Clark, J. S., Salk, C., Melillo, J., & Mohan, J. (2014). Tree phenology responses to winter chilling, spring warming, at north and south range limits. *Functional Ecology*, 28, 1344–1355. <https://doi.org/10.1111/1365-2435.12309>
- Črepinšek, Z., Kajfež-Bogataj, L., & Bergant, K. (2006). Modelling of weather variability effect on fitophenology. *Ecological Modelling*, 194, 256–265. <https://doi.org/10.1016/j.ecolmodel.2005.10.020>
- De Reaumur, R.-A. (1735). Observations du thermometre, faites a Paris pendant l'annee 1735 comparees avec celles qui onte faites sous la ligne et al'Ile de France, a Alger et en quelques-unes de nos iles de l'Amerique. *Memoires de l'Academie Royale des Sciences de Paris*, 1735, 545–576.
- Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., ... Vitart, F. (2011). The ERA-interim reanalysis: configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, 137, 553–597. <https://doi.org/10.1002/qj.828>
- Fisher, J. I., Richardson, A. D., & Mustard, J. F. (2007). Phenology model from surface meteorology does not capture satellite-based greenup estimations. *Global Change Biology*, 13, 707–721. <https://doi.org/10.1111/j.1365-2486.2006.01311.x>
- Friedl, M. A., Sulla-Menashe, D., Tan, B., Schneider, A., Ramankutty, N., Sibley, A., & Huang, X. M. (2010). MODIS collection 5 global land cover: algorithm refinements and characterization of new datasets. *Remote Sensing of Environment*, 114, 168–182. <https://doi.org/10.1016/j.rse.2009.08.016>
- García-Mozo, H., Galán, C., Belmonte, J., Bermejo, D., Candau, P., Díaz de la Guardia, C., ... Chuine, I. (2009). Predicting the start and peak dates of the Poaceae pollen season in Spain using process-based models. *Agricultural and Forest Meteorology*, 149, 256–262. <https://doi.org/10.1016/j.agrformet.2008.08.013>
- Gill, A. L., Gallinat, A. S., Sanders-DeMott, R., Rigden, A. J., Short Gianotti, D. J., Mantooth, J. A., & Templer, P. H. (2015). Changes in autumn senescence in northern hemisphere deciduous trees: A meta-analysis of autumn phenology studies. *Annals of Botany*, 116, 875–888. <https://doi.org/10.1093/aob/mcv055>
- Hänninen, H. (1990). Modelling bud dormancy release in trees from cool and temperate regions. *Acta Forestalia Fennica*, 213, 1–47. <https://doi.org/10.14214/aff.7660>
- Hänninen, H., & Kramer, K. (2007). A framework for modelling the annual cycle of trees in boreal and temperate regions. *Silva Fennica*, 41, 167–205. <https://doi.org/10.14214/sf.313>
- Haylock, M. R., Hofstra, N., Klein Tank, A. M. G., Klok, E. J., Jones, P. D., & New, M. (2008). A European daily high-resolution gridded data set of surface temperature and precipitation for 1950–2006. *Journal of Geophysical Research Atmospheres*, 113, D20119. <https://doi.org/10.1029/2008JD01020>
- Hollinger, D. Y., Goltz, S. M., Davidson, E. A., Lee, J. T., Tu, K., & Valentine, H. T. (1999). Seasonal patterns and environmental control of carbon dioxide and water vapour exchange in an ecotonal boreal forest. *Global Change Biology*, 5, 891–902. <https://doi.org/10.1046/j.1365-2486.1999.00281.x>
- Hufkens, K., Keenan, T. F., Flanagan, L. B., Scott, R. L., Bernacchi, C. J., Joo, E., ... Richardson, A. D. (2016). Productivity of North American grasslands is increased under future climate scenarios despite rising aridity. *Nature Climate Change*, 6, 10–14.
- Hunter, A. F., & Lechowicz, M. J. (1992). Predicting the timing of budburst in temperate trees. *Journal of Applied Ecology*, 29, 597–604. <https://doi.org/10.2307/2404467>
- Jeong, S., & Medvigy, D. (2014). Macroscale prediction of autumn leaf coloration throughout the continental United States. *Global Ecology and Biogeography*, 23, 1245–1254. <https://doi.org/10.1111/geb.12206>
- Keenan, T., Darby, B., Felts, E., Sonnentag, O., Friedl, M., Hufkens, K., ... Richardson, A. D. (2014). Tracking forest phenology and seasonal physiology using digital repeat photography: A critical assessment. *Ecological Applications*, 24, 1478–1489. <https://doi.org/10.1890/13-0652.1>
- Klosterman, S. T., Hufkens, K., Gray, J. M., Melaas, E., Sonnentag, O., Lavine, I., ... Richardson, A. D. (2014). Evaluating remote sensing of deciduous forest phenology at multiple spatial scales using PhenoCam imagery. *Biogeosciences Discussions*, 11, 2305–2342. <https://doi.org/10.5194/bgd-11-2305-2014>
- Kramer, K. (1994). Selecting a model to predict the onset of growth of *Fagus-sylvatica*. *Journal of Applied Ecology*, 31, 172–181. <https://doi.org/10.2307/2404609>
- Landsberg, J. J. (1974). Apple fruit bud development and growth; analysis and an empirical model. *Annals of Botany*, 38, 1013–1023. <https://doi.org/10.1093/oxfordjournals.aob.a084891>
- Lang, G., Early, J., Martin, G., & Darnell, R. (1987). Endo-, para-, and eco-dormancy: Physiological terminology and classification for dormancy research. *HortScience*, 22, 371–377.
- Laube, J., Sparks, T. H., Estrella, N., Höfler, J., Ankerst, D. P., & Menzel, A. (2013). Chilling outweighs photoperiod in preventing precocious spring development. *Global Change Biology*, 20, 170–182.
- Laube, J., Sparks, T. H., Estrella, N., & Menzel, A. (2014). Does humidity trigger tree phenology? Proposal for an air humidity based framework for bud development in spring. *New Phytologist*, 202, 350–355. <https://doi.org/10.1111/nph.12680>

- Leinonen, I., Repo, T., & Hänninen, H. (1997). Changing environmental effects on frost hardness of scots pine during dehardening. *Annals of Botany*, 79, 133–137. <https://doi.org/10.1006/anbo.1996.0321>
- Linkosalo, T., Häkkinen, R., & Hänninen, H. (2006). Models of the spring phenology of boreal and temperate trees: Is there something missing? *Tree physiology*, 26, 1165–1172. <https://doi.org/10.1093/treephys/26.9.1165>
- Masle, J., Doussinault, G., Farquhar, G. D., & Sun, B. (1989). Foliar stage in wheat correlates better to photothermal time than to thermal time. *Plant, Cell & Environment*, 12, 235–247. <https://doi.org/10.1111/j.1365-3040.1989.tb01938.x>
- Mebane, W. R. J., & Sekhon, J. S. (2011). Genetic optimization using derivatives: The rgenoud package for R. *Journal of Statistical Software*, 42, 1–26.
- Melaas, E. K., Friedl, M. A., & Richardson, A. D. (2016). Multiscale modeling of spring phenology across deciduous forests in the Eastern United States. *Global Change Biology*, 22, 792–805. <https://doi.org/10.1111/gcb.13122>
- Migliavacca, M., Sonnentag, O., Keenan, T. F., Cescatti, A., O'Keefe, J., & Richardson, A. D. (2012). On the uncertainty of phenological responses to climate change, and implications for a terrestrial biosphere model. *Biogeosciences*, 9, 2063–2083. <https://doi.org/10.5194/bg-9-2063-2012>
- Murray, M. B., Cannell, M. G. R., & Smith, R. I. (1989). Date of budburst of fifteen tree species in Britain following climatic warming. *Journal of Applied Ecology*, 26, 693–700. <https://doi.org/10.2307/2404093>
- R Core Team (2016). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Richardson, A. D., Anderson, R. S., AltafArain, M., Barr, A. G., Bohrer, G., Chen, G., ... Xue, Y. (2011). Terrestrial biosphere models need better representation of vegetation phenology: Results from the North American Carbon Program. *Global Change Biology*, 18, 566–584. <https://doi.org/10.1111/j.1365-2486.2011.02562.x>
- Richardson, A. D., Black, T. A., Ciais, P., Delbart, N., Friedl, M. A., Gobron, N., ... Varlagin, A. (2010). Influence of spring and autumn phenological transitions on forest ecosystem productivity. *Philosophical transactions of the Royal Society of London Series B, Biological Sciences*, 365, 3227–3246. <https://doi.org/10.1098/rstb.2010.0102>
- Richardson, A. D., Hufkens, K., Milliman, T., Aubrecht, D. M., Chen, M., Gray, J. M., ... Frolking, S. (2018). Tracking vegetation phenology across diverse North American biomes using PhenoCam imagery. *Scientific Data* (in press).
- 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. *Agricultural and Forest Meteorology*, 169, 156–173. <https://doi.org/10.1016/j.agrformet.2012.09.012>
- Rohde, R., Muller, R., Jacobsen, R., Muller, E., Groom, D., & Wickham, C. (2012). A New estimate of the average earth surface land temperature spanning 1753 to 2011. *Geoinformatic & Geostatistics: An Overview*, 1, 1–7.
- Rosenzweig, C., Casassa, G., Karoly, D.J., Imeson, A., Liu, C., Menzel, A., ... Tryjanowski, P. (2007). Assessment of observed changes and responses in natural and managed systems. In M. L. Parry, O. F. Canziani, J. P. Palutikof, P. J. van der Linden & C. E. Hanson (Eds.), *Climate Change 2007: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. (pp. 79–131). Cambridge, UK: Cambridge University Press.
- Sakai, R. K., Fitzjarrald, D. R., & Moore, K. E. (1997). Detecting leaf area and surface resistance during transition seasons. *Agricultural and Forest Meteorology*, 84, 273–284. [https://doi.org/10.1016/S0168-1923\(96\)02359-3](https://doi.org/10.1016/S0168-1923(96)02359-3)
- Schaber, J., & Badeck, F.-W. (2003). Physiology-based phenology models for forest tree species in Germany. *International journal of biometeorology*, 47, 193–201. <https://doi.org/10.1007/s00484-003-0171-5>
- Sonnentag, O., Hufkens, K., Teshera-Sterne, C., Young, A. M., Friedl, M., Braswell, B. H., ... Richardson, A. D. (2012). Digital repeat photography for phenological research in forest ecosystems. *Agricultural and Forest Meteorology*, 152, 159–177. <https://doi.org/10.1016/j.agrformet.2011.09.009>
- Thornton, P.E., Thornton, M.M., Mayer, B.W., Wilhelmi, N., Wei, Y., Devarakonda, R., & Cook, R.B. (2017). Daymet: Daily Surface Weather Data on a 1-km Grid for North America, Version 3. Data set. Oak Ridge National Laboratory Distributed Active Archive Center, Oak Ridge, Tennessee, USA. Retrieved from: https://daac.ornl.gov/cgi-bin/ds-viewer.pl?ds_id=1328
- Toomey, M., Friedl, M. A., Frolking, S., Hufkens, K., Klosterman, S., Sonnentag, O., ... Richardson, A. D. (2015). Greenness indices from digital cameras predict the timing and seasonal dynamics of canopy-scale photosynthesis. *Ecological Applications*, 25, 99–115. <https://doi.org/10.1890/14-0005.1>
- Tsallis, C., & Stariolo, D. A. (1996). Generalized Simulated Annealing. *Physica A*, 233, 395–406. [https://doi.org/10.1016/S0378-4371\(96\)00271-3](https://doi.org/10.1016/S0378-4371(96)00271-3)
- Tuck, S. L., Phillips, H. R. P., Hintzen, R. E., Scharlemann, J. P. W., Purvis, A., & Hudson, L. N. (2014). MODISTools - downloading and processing MODIS remotely sensed data in R. *Ecology and Evolution*, 4, 4658–4668. <https://doi.org/10.1002/ece3.1273>
- Vitasse, Y., Porté, A. J., Kremer, A., Michalet, R., & Delzon, S. (2009). Responses of canopy duration to temperature changes in four temperate tree species: Relative contributions of spring and autumn leaf phenology. *Oecologia*, 161, 187–198. <https://doi.org/10.1007/s00442-009-1363-4>
- Wang, J. Y. (1960). A critique of the heat unit approach to plant response studies. *Ecology*, 41, 785–790. <https://doi.org/10.2307/1931815>
- White, M. A., De Beurs, K. M., Didan, K., Inouye, D. W., Richardson, A. D., Jensen, O. P., ... Lauenroth, W. K. (2009). Intercomparison, interpretation, and assessment of spring phenology in North America estimated from remote sensing for 1982–2006. *Global Change Biology*, 15, 2335–2359. <https://doi.org/10.1111/j.1365-2486.2009.01910.x>
- Xiang, Y., Gubian, S., Suomela, B., & Hoeng, J. (2013). Generalized simulated annealing for global optimization: The GenSA Package. *R Journal*, 5, 13–28.
- Xin, Q., Broich, M., Zhu, P., & Gong, P. (2015). Modeling grassland spring onset across the Western United States using climate variables and MODIS-derived phenology metrics. *Remote Sensing of Environment*, 161, 63–77. <https://doi.org/10.1016/j.rse.2015.02.003>
- Zhang, X., Friedl, M. A., Schaaf, C. B., Strahler, A. H., Hodges, J. C. F., Gao, F., ... Huete, A. (2003). Monitoring vegetation phenology using MODIS. *Remote Sensing of Environment*, 84, 471–475. [https://doi.org/10.1016/S0034-4257\(02\)00135-9](https://doi.org/10.1016/S0034-4257(02)00135-9)

SUPPORTING INFORMATION

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

How to cite this article: Hufkens K, Basler D, Milliman T, Melaas EK, Richardson AD. An integrated phenology modelling framework in R. *Methods Ecol Evol*. 2018;9:1276–1285. <https://doi.org/10.1111/2041-210X.12970>