https://doi.org/10.5194/egusphere-2026-1710
Preprint. Discussion started: 8 April 2026

The growing seasons of global forest ecosystems from 1850 to 2100 estimated with a probabilistic temperature-based model

Pierluigi R. Guaita[1,2], Giacomo Gerosa[1], Riccardo Marzuoli[1]

[1]Department of Mathematics and Physics, Catholic University of the Sacred Heart, Brescia, Italy
[2]Department of Applied Computational Mathematics and Statistics, University of Notre Dame, Notre Dame, IN, USA

Correspondence to: pierluigirenan.guaita@unicatt.it, giacomo.gerosa@unicatt.it

Abstract

Global climate warming has significantly altered forest phenology in the past decades, with measurable shifts in the timing and duration of the growing season (GS). These changes are expected to intensify in the future, potentially affecting both ecosystem productivity and land-atmosphere interactions. Accurately representing GS dynamics is therefore essential for assessing ecosystem vulnerability and improving the representation of vegetation processes in Earth system models. Here, we introduce GS-P, a probabilistic, temperature-based model developed within a machine-learning framework to estimate the start and end of the growing season (SGS and EGS) for global forest ecosystems over the period 1850-2100.

Results show stable GS timing until the 1970s, followed by significant shifts characterized by earlier SGS and later EGS, leading to a global extension of the GS. Under future climate scenarios, GS duration is projected to increase by approximately one month under low-emission conditions and up to two months under high-emission scenarios, with stronger responses in the Northern Hemisphere. Compared to alternative models, GS-P achieves comparable or improved predictive accuracy while exhibiting greater extrapolation capabilities and providing explicit uncertainty estimates. Furthermore, the model effectively represents key ecological features, such as stronger temperature control and greater spatial heterogeneity in spring than autumn phenology, and detection of regions where temperature alone provides limited explanatory power, suggesting a stronger role of additional drivers. Additionally, GS-P enables the identification of regions characterized by transitional states and high prediction uncertainty, potentially reflecting climate–ecosystem disequilibrium and enhanced ecosystem vulnerability. This model provides a flexible and interpretable framework for simulating GS dynamics at the global scale, offering improved constraints for carbon cycle modelling and supporting the assessment of ecosystem responses to future climate change.

1 Introduction

Vegetation seasonal cycles are a fundamental component of the interactions between ecosystems and environment and are essential in regulating Earth's climate system (Bonan 2016). At a biogeophysical level, vegetation growing phases and the intensity of its biological activity can affect surface albedo, roughness length and evapotranspiration, altering local water

https://doi.org/10.5194/egusphere-2026-1710
Preprint. Discussion started: 8 April 2026

budget and surface energy balance (Miralles et al. 2025). At the biogeochemical level, plants are central to the global carbon cycle, influencing planetary radiative balance (Stuart Chapin III et al. 2009), to the cycling of other key nutrients, such as nitrogen and phosphorous (Fowler et al. 2013; Buendía et al. 2010) and to the exchange of reactive trace gases; specifically, vegetation represents a critical sink for tropospheric ozone and modulates atmospheric oxidative capacity (Emberson 2020). Conversely, environmental variables such as air temperature, radiation, and water availability, interact with plant physiological controls (e.g. genetics, hormones, metabolism) to constrain vegetation growth at species-specific level (Chapin et al. 2011).

Understanding and accurately predicting the *growing season* (GS), i.e. the period of plant biological activity, is therefore essential for projecting future ecosystem-climate feedbacks under future global changes. Körner et al. (2023) recently proposed four definitions for GS, each capturing specific biological processes and vegetation-environment interactions: the GS *sensu stricto*, the *phenological season*, the *productive season*, and the *meteorological season*. Each definition offers unique insights with its own advantages and drawbacks, but their practical applicability largely depends on which aspect are covered when referring to the term GS. The GS *sensu stricto* refers to actual tissue growth. However, these processes are difficult to measure, as they do not necessarily correspond to more visible developmental stages. On the other hand, the *phenological season* is defined by visible markers such as leaf flushing (green-up) and senescence (green-down). These markers are more readily monitored via phenological observations and remote sensing, but they do not necessarily reflect actual biomass accumulation and growth. The *productive season* relies on productivity metrics like net biomass increase, net primary production (NPP) or net ecosystem exchange (NEE). While NEE can be quantified via eddy covariance measurements of CO_2 fluxes, net biomass increase has no clear-cut seasonal definition, and NPP (canopy photosynthesis minus autotrophic respiration) remains difficult to measure in real-world ecosystems. The *meteorological season* relies on environmental thresholds (typically air temperature or soil water availability to the plant) to determine the GS, following the notion that long-term climate dynamics induce evolutionary pressure on plants, broadly determining their physiological mechanisms. However, growth and developmental stages could be modified or halted by adverse meteorological events, and the meteorological season, in principle, cannot account for such occurrences. As such, the meteorological season strictly represents the climatic conditions for potential vegetation activity.

It is apparent that the meteorological and the phenological definitions of GS are more practical than the others, despite their respective caveats. The meteorological season is frequently used to estimate the GS, due to its robustness in describing long-term shifts in developmental stages and biological activity (Y. Mo et al. 2023). The phenological season is particularly effective for calibration and validation of ecosystem and climate models, because it is based on direct observations of phenological markers (e.g., Hufkens et al. 2018; Zheng et al. 2022). Satellite-based remote sensing has become increasingly important in this regard in the last decades, offering global coverage, scalability of the results and broad applicability (Zhang et al. 2003; Friedl et al. 2010; Melaas et al. 2013; Claverie et al. 2018). These tools detect phenological events by analyzing reflectance properties of vegetation; in fact, active plants tend to absorb a significant amount of red light (620-670 nm) and

https://doi.org/10.5194/egusphere-2026-1710
Preprint. Discussion started: 8 April 2026

reflect near-infrared light (750-900 nm) due to leaf structure and chlorophyll content. These spectral properties are synthesized into indexes such as the Normalized Difference Vegetation Index (NDVI) to determine key phenological dates.

While process-based models can predict phenological stages by simulating plant physiological responses to environmental conditions, they are often more computationally intensive than their statistical counterpart and require detailed parameterization, with uncertainties that are difficult to quantify (Chuine et al. 2013; Meier and Bigler 2023; Garnot et al. 2025). Alternatively, several studies frequently linked meteorological data to phenological observations to estimate GS using empirical models and often validating model outputs against ground-based phenological records or satellite-derived indices (Y. Mo et al. 2023). However, these models are often limited to specific geographical regions, or to individual vegetation species (e.g., Linderholm et al. 2008; Blümel and Chmielewski 2012).

The primary task of this study is to determine the Start and End of the Growing Season (SGS and EGS) for generic forest ecosystems at a global scale, using a novel probabilistic, temperature-based model (GS-P) developed within a machine-learning framework. To assess the model's performance and robustness, we conduct a comparative analysis against established deterministic models and a latitudinal benchmark (GS-Lat), using ERA5 reanalysis and MODIS remote-sensing data for calibration and validation. The ultimate outcome is the creation of an open-source dataset that reconstructs, with quantified uncertainty, the trajectories of SGS and EGS from 1850 to 2100 under different climate change scenarios.

Several reasons motivate this effort. First, such a product can provide continuous and consistent phenological estimates for forest ecosystems, allowing us to infer changes in phenological transitions during the considered two and a half centuries. This product may help reduce uncertainties in terrestrial carbon cycle models and could support adaptive forest management strategies in a changing climate. Second, we seek to improve the capability of phenological modelling in predicting the SGS and EGS, and to assess the potential advantages of machine-learning approaches over more traditional methods. Namely, we aim to overcome previous limitations in spatial extent, species specificity, and extrapolation skills under climate change scenarios, by providing a robust, generalizable framework for prediction of phenological stages. Third, using temperature as the sole predictor allows us to clarify its role as a driver of phenological transitions at global scale. In this context, model uncertainty is not only a measure of predictive performance but may also help to identify regions where additional controls on phenological timing (e.g., precipitation seasonality) are likely to be more important than temperature alone, and where future climate change may influence forest phenology and, more broadly, forest ecosystem functioning.

2 Data and methods

The study is performed using GS-P, a probabilistic temperature-based model (i.e., a boosted decision tree logistic classifier) specifically developed to predict phenological stages of forest ecosystems at a global scale. Two more traditional temperature-based models (GS-Lin2, Linderholm et al. 2008; GS-BC2, Blümel and Chmielewski 2012) are employed for comparison, and a latitudinal model (GS-Lat, Zhang et al. 2004; Simpson et al. 2012) is used for benchmarking.

https://doi.org/10.5194/egusphere-2026-1710
Preprint. Discussion started: 8 April 2026

2.1 Data

Calibration of the considered four GS models is performed by linking ERA5 (Hersbach et al. 2020) reanalysis temperature products with products from the remote-sensing MODIS land data suite (MCD12Q1, MCD12Q2; Friedl et al. 2010; Zhang et al. 2003). Simulated temperature fields from UKESM1-0-LL (Sellar et al. 2019) are used to predict the growing season in past and future periods.

2.1.1 Temperature data

ERA5 is the fifth generation of atmospheric reanalysis produced by the European Centre for Medium-Range Weather Forecasts under the Copernicus Climate Change Service. By assimilating data from observations satellites, ground stations, and other sources, it provides hourly estimates of atmospheric, land-surface, and ocean-wave variables at a spatial resolution of approximately 31 km globally, spanning from 1950 to the present. For this study, we use the daily 2-meter surface air temperature.

UKESM1 is a fully coupled earth system model (ESM) developed by the UK Met Office, which contributes to the Coupled Model Intercomparison Project Phase 6 (CMIP6; Eyring et al. 2016) and provides global climate simulations for both the historical time and under future climate change scenarios at a $1.25° \times 1.875°$ spatial resolution. This study considers the *historical* experiment (Eyring et al. 2016), which is designed to simulate climate from 1850 to 2014 using historical forcings. For future times (2015-2100), we primarly consider SSP1-2.6 (Van Vuuren et al. 2017) and SSP5-8.5 (Kriegler et al. 2017), a low- and high-radiative forcing scenario, respectively. Furthermore, the SSP3-7.0 (Fujimori et al. 2017) is also considered: the results from this scenario are not presented in maps and are not discussed in this study (as they are very similar to the ones from SSP5-8.5). However, values related to this scenario may be found in tables, and the associated GS-P estimations for SGS and EGS may be found in the final published dataset. For all experiments, the run variant r1i1p1f2 is used (Tang et al. 2019; Good et al. 2019a, 2019b, 2019c).

2.1.2 Phenological data

We used the MCD12Q1 and MCD12Q2 products from the MODIS land data suite to retrieve land cover and phenology data for all available years (from 2001 to 2023, at the time of data preprocessing) at a 500-meter resolution. More specifically, the FAO-Land Cover classification system from MCD12Q1 is used to select all areas with more than 60% forest coverage. Over the nodes selected this way, the MCD12Q2 product provides four fields that describe phenology phases: greenup, mid-greenup, mid-greendown, and dormancy. These fields are identified via the 2-band Enhanced Vegetation Index (EVI-2; D. Mo et al. 2012), similar to the Normalized Difference Vegetation Index, but designed to improve sensitivity to high-biomass regions, reduce aerosols disturbances, and minimize soil background effects. Each phenological field is filtered on the chosen land use and, after being checked for quality assurance/quality control, upscaled at the ERA5 resolution by averaging

https://doi.org/10.5194/egusphere-2026-1710
Preprint. Discussion started: 8 April 2026

all MODIS nodes within each ERA5 node. For each year, we identified and removed the outliers, i.e., the values whose deviation from the local spatial mean exceeded twice the standard deviation of the MODIS fields of the year considered.

2.2 Description of GS models

This study selected GS models that (1) are based mostly on temperature, and that (2) have simple criteria to determine the SGS and EGS. Two models are based on the ones presented by Linderholm et al. (2008), and Blümel and Chmielewski (2012). These models were originally tailored either to specific species, to limited geographical regions, or to reproduce specific phenological stages. In this study, their scope is expanded to include generic forests ecosystems with a global coverage, and for both SGS and EGS. As an updated version of the original GS models, they are hereby named after their creators (GS-Lin2, GS-BC2). Then, the developed probabilistic temperature-based model (GS-P), based on the estimation of the likelihood of having a GS at a given location and time, is described. Finally, a simple latitude model (GS-Lat) that does not include any explicit relationship with temperature is also included for benchmarking purposes.

2.2.1 GS-Lin2

Linderholm et al. (2008) employed simple criteria based on warm spells, cold spells and temperature thresholds to determine the start and the end of the meteorological GS across the Greater Baltic Area. Four parameters determine the SGS and the EGS within the GS-Lin2 model. Namely, the SGS is defined as the last day of the first warm spell, i.e. the last day of a series of n_w days with daily temperatures exceeding a temperature threshold of T_w. Similarly, the EGS is defined as the first day of the first cold spell (n_c subsequent days) with daily temperatures below a temperature threshold of T_c. As these criteria are not very restrictive by themselves and could lead to identifying multiple short GSs in the same year, GS-Lin2 is set to prevent SGS from occurring over the Northern (Southern) Hemisphere in Oct-Dec (Apr-Jun), and the EGS occurring in Apr-Jun (Oct-Dec). Furthermore, during a certain year, the SGS is required to start before the EGS in the Northern Hemisphere (NH), while in the Southern Hemisphere (SH) this requirement is the other way around.

2.2.2 GS-BC2

The Blümel and Chmielewski (2012) model was originally formulated to calculate the beginning of apple blossom in Germany, building on the notion that heat accumulation releases ecodormancy in plants, thus acting as forcing. Compared to other models, an additional feature is that the heat accumulation is adjusted on the photoperiod, and in this sense this model is not purely temperature based. This model was selected because it was reported to perform well comparatively to others, for both deciduous needleleaf forest and for mixed forests Y. Mo et al. (2023).

For GS-BC2, the forcing at a given timestep t ($F_{SGS}(t)$; [°C d]) for the SGS is defined through the following formula:

https://doi.org/10.5194/egusphere-2026-1710
Preprint. Discussion started: 8 April 2026

$$F_{SGS}(t) = \sum_{i=t_{1,SGS}}^{t} \max\left(0, T_i - T_{BF,SGS}\right) \cdot \Delta t \cdot \left(\frac{L_i}{10}\right)^{a_{SGS}} \quad (1)$$

Where $t_{1,SGS}$ is the start of the cumulation ([d]) either from the 1st of January for the NH or from the 1st of July for the SH, T_i is the daily mean temperature ([°C]), $T_{BF,SGS}$ is the base temperature for heat accumulation ([°C]), Δt is the timestep considered for the summation (in this case, $\Delta t = 1$ d), L_i is the photoperiod ([h]), which is normalized over 10 hours, and a_{SGS} is a parameter to modulate the effect of the photoperiod. The exceedance of a certain critical threshold $F_{crit,SGS}$ by the forcing $F(t)$ at a certain timestep determines the timing of the SGS (t_{SGS}). Therefore, this model requires four parameters ($t_{1,SGS}$, $F_{crit,SGS}$, $T_{BF,SGS}$, and a_{SGS}).

The usage on SGS is expanded also to the EGS, by substituting the maximum in Eq. (1) with a minimum (i.e., cold accumulation), and by referring new parameters to the EGS ($t_{1,EGS}$, $F_{crit,EGS}$, $T_{BF,EGS}$, and a_{EGS}). In this case, $t_{1,EGS}$ is the number of days from the 1st of July for the NH, and from the 1st of January for the SH. It should be noted that using a minimum instead of a maximum, allows the cumulation of negative quantities, and therefore the $F_{crit,EGS}$ has negative values. As cold cumulation should matter more the shorter is the photoperiod, a_{EGS} is also set to be negative.

2.2.3 GS-P

The probabilistic model GS-P estimates the SGS and EGS from temperature dynamics. It relies on the following temperature-based predictors: the 14-day moving average of the daily temperatures (T_{14d}), the associated regression slope (∇T_{14d}) and moving standard deviation (σ_{14d}), and the 60-day moving average (T_{60d}). This allows to capture not only the fact that GS is determined by the current short- and long-term temperature averages, but also how stability, and warming and cooling trends can signal transitions into and out the GS.

More formally, the GS-P model is a two-step model. First, a binary classification model uses $T_{14d}, \nabla T_{14d}, \sigma_{14d}, T_{60d}$ as input to determine the daily likelihood of having or not having a GS (0=non-GS, 1=GS). Then, the SGS and the EGS are identified by defining probability thresholds p_{SGS} and p_{EGS} that mark the transitions between GS and non-GS. However, transitions between GS and non-GS defined this way can occur in principle multiple times during a year and can also happen within a short time frame. For this reason, the GS-P model identifies the SGS and the EGS as the first and the last of the respective transition days. Further details on the GS-P model are available in Appendix A.1.

2.2.4 GS-Lat

The latitudinal model has been used to determine the general SGS and EGS over the NH e.g. in the framework of the LRTAP convention and in the EMEP model (Simpson et al. 2012; LRTAP Convention 2017). In the formulation therein, the SGS and EGS times (t_{SGS} and t_{EGS}) are determined as follows:

$$t_{SGS} = 105 + 1.5 \cdot (l - 50) + 10 \cdot e$$

$$t_{EGS} = 297 - 2 \cdot (l - 50) - 10 \cdot e$$

Where l is the latitude ([°]) and e is the elevation ([km]).

https://doi.org/10.5194/egusphere-2026-1710
Preprint. Discussion started: 8 April 2026

As this model, at its core, is a linear regression with l and e as covariates, we define GS-Lat for the SGS as:

$$t_{SGS} = \beta_0 + \beta_1|l| + \beta_2 e + \varepsilon \quad (2)$$

Where β_i $(i = 0,1,2)$ the coefficients, and $\varepsilon \in N(0, \sigma^2)$ the errors. The t_{EGS} is determined similarly using corresponding coefficients. Furthermore, GS-Lat is calibrated separately for SH and NH, leading to two linear regressions for SGS and two for EGS.

2.3 Parameters calibration

The MODIS product is used for calibration and validation of the models, hence linking the SGS and EGS inferred from EVI-2 (phenological season) with temperatures (meteorological season) from ERA5. Of the years that can be used for training the GS models (2001-2023), a randomly selected 75% of the observations, together with the complete years 2005, 2010, 2015, and 2020 are used for calibration. The remaining 25%, alongside with years 2006, 2011, 2016, and 2021 are kept for validation.

GS-Lin2, GS-BC2 and GS-P all rely on temperature-related parameterizations. However, reasonable parameterizations should depend on the climate type of each region of the globe, which broadly determines phenological transitions behaviour. For this reason, we define regions based on the Köppen-Geiger (KG) climate classification (Peel et al. 2007) and calibrate a different set of parameters over each of these regions. The KG classification identifies each climate through three letters: the first indicating the main group, the second precipitation seasonality, and the third temperature seasonality. In this study, the main groups B (dry climate) and E (polar climate) are excluded due to the absence of forests, while the group A (tropical climate) was not considered because the meteorological GS for humid tropical forests is all year round. Here, we calibrate the three temperature-based models over each subgroup of the temperate climate (C) and of the continental climate (D) accounting only for temperature seasonality. For this reason, the second letter of the KG classification is omitted from the notation (e.g. “C*a” indicates temperate climate with hot summer regardless of seasonal precipitation). Furthermore, since C*c has very few occurrences around the globe, it was aggregated to C*b. Criteria to define each region are listed in Table S1. Figure S1 and Fig. S2 show global maps of KG classes obtained from ERA5 and UKESM1-0-LL temperatures, respectively.

For GS-Lin2 and GS-BC2, the modelled SGS and EGS are evaluated against satellite observations, and the parameters are estimated from the parameter combination that leads to the lowest error (i.e., that minimizes the loss function). The loss function is based on the Root Mean Squared Error (RMSE) of the SGS and EGS. However, for given nodes and years, the temperature-based models fail to predict transitions when temperatures that should allow to identify the GS never satisfy the given criteria for GS to start or end. To correct this feature and favour a calibration of the parameters that can effectively predict SGS and EGS, the loss function includes a penalty proportional to the fraction of missing predictions. For each KG classification, the parameters of GS-Lin2 and GS-BC2 are calibrated by simulated annealing (a probabilistic optimization algorithm for search of local minima; Delahaye et al. 2019), using 3000 iterations, and frequent reannealing (every 50 iterations).

https://doi.org/10.5194/egusphere-2026-1710
Preprint. Discussion started: 8 April 2026

The GS-Lat model, being a linear regression based purely on geographical coordinates, is calibrated globally via ordinary least squares, without accounting for KG classes.

In GS-P, the binary classification model is constituted by an ensemble of decision trees, calibrated with the XGBoost algorithm (Chen and Guestrin 2016). For this part of GS-P, there are no parameters in the usual statistical sense, but the behaviour of the classifier can be understood by considering the probability profiles with respect to each of the input variables ($x = [T_{14d}, \nabla T_{14d}, \sigma_{14d}, T_{60d}]$; Fig. S3). In each grid node, the probability thresholds that determine GS transitions can be determined by first associating observed SGS and EGS dates from MODIS with the corresponding GS probabilities. Then, for each KG class, we perform node-specific parameter inference via empirical Bayesian estimation under a hierarchical Normal-Normal model using a shrinkage estimator (Gelman et al. 2025). This hierarchical model assumes that each node-specific probability threshold is drawn from a population distribution (prior) whose parameters (mean μ, between-node variance τ^2, within-node variance σ^2) are empirically estimated from observations. The node-specific p_{SGS} and p_{EGS} (posterior means) are estimated as weighted averages of the node-specific sample mean and the global mean, with weights reflecting the relative uncertainty of each node's observations (shrinkage). For more details on parameter calibration of GS models, see Appendix A.2.

For the GS-Lat model, parameters' confidence intervals (CIs) are provided directly by the regression models. However, for the parameters of GS-Lin2 and GS-BC2, traditional CIs for the parameters cannot be derived, because the parameters have a non-linear structure and inference on them is performed via a global optimization heuristic (simulated annealing). Typical CI estimation in this case would involve non-parametric methods (e.g., bootstrapping), but they would be unpractical due to computational costs. For the Bayesian estimator hyperparameters and for the probability thresholds of GS-P, CIs are determined via non-parametric bootstrapping over the years (resampling entire maps).

2.4 Performance evaluation and prediction uncertainty

The evaluation of the models is performed in terms of Mean Bias (MB), Mean Absolute Error (MAE), and Pearson's correlation (ρ) using the validation dataset and separately for SGS and EGS. The MB is defined as:

$$\mathrm{MB} = \frac{\sum_{i \in I} \hat{Y}_i - Y_i}{n} \quad (3)$$

Where Y_i are the MODIS observation, $\hat{Y}_i$ are the values predicted by the model, and $I = \{1, \dots, n\}$ is the number of values for which the model actually predicts dates for SGS and EGS. The MAE is defined as:

$$\mathrm{MAE} = \frac{\sum_{i \in I} |\hat{Y}_i - Y_i|}{n} \quad (4)$$

The ρ is defined as:

$$\rho = \frac{\mathrm{Cov}(Y, \hat{Y})}{\sigma_Y \sigma_{\hat{Y}}} \quad (5)$$

However, these measures do not account for the number of times $\tilde{n}$ in which the models failed in providing actual predictions. Therefore, MB and MAE are corrected akin to the loss function λ, multiplying their values by the same

https://doi.org/10.5194/egusphere-2026-1710
Preprint. Discussion started: 8 April 2026

240 penalization factor based on the number of failed predictions (Eq. A2). ρ is also corrected similarly, but in this case the metric is divided by the penalization factor. Performance measures are calculated over individual KG classes and globally.

To compare the performance the different GS model, we compute skill scores for MAE and ρ following (Murphy 1988), using the GS-Lat model for benchmarking. The skill score (SS) for a given metric A expresses the improvement achieved by a model GS-X over the benchmarking model GS-Lat as a fraction (in percent) of the difference in the metric between a

245 hypothetical perfect model (GS-PM) and GS-Lat. Positive values indicate improvement over GS-Lat, while negative values indicate degradation:

$$SS(\text{GS-X}) = 100 \cdot \frac{A(\text{GS-X}) - A(\text{GS-Lat})}{A(\text{GS-PM}) - A(\text{GS-Lat})} \quad (6)$$

In this case, $\text{MAE}(\text{GS-PM}) = 0$ and $\rho(\text{GS-PM}) = 1$.

Uncertainty in the predictions is quantified using spread functions on bootstrapped absolute errors. For each observation, we pair the GS model prediction with its absolute error, forming the sample set $\left(\hat{Y}_i, \left|\hat{Y}_i - Y_i\right|\right)_{i \in I}$. This set is then resampled

250 5,000 times (bootstrapping) and, for each bootstrap sample, we fit a second-degree polynomial (spread function). To estimate the uncertainty associated with a given prediction value $\hat{Y}$, the 5,000 spread functions are evaluated in $\hat{Y}$. The 95th percentile of the resulting distribution of error estimates (p_{95}) provides the upper bound of the uncertainty range, leading to a prediction interval of the form $\hat{Y} \pm p_{95}$. Uncertainty in the yearly hemispherical SGS and EGS averages (Figure 3) is quantified using a Monte Carlo framework that propagates model error: spatially and temporally correlated perturbations,

255 scaled by bootstrapped error spread functions, are added to the predictions and hemispherical means are recomputed for each realization. The final hemispherical estimate is given by the ensemble median, while the 2.5th and 97.5th percentiles define the uncertainty bounds, representing the expected range of hemispherical SGS and EGS values under model uncertainty.

3 Results

3.1 Parameters spaces

260 The disrupted relationship between phenological transitions, latitude and elevation is also implied by the wider 95% CIs in the SH, especially for the coefficient related to elevation (Table S2; SGS: [-8.97,-7.94]; EGS: [-6.29,-5.42]).

https://doi.org/10.5194/egusphere-2026-1710
Preprint. Discussion started: 8 April 2026

Table 1 lists all inferred parameters for each GS model, separated by KG class, and with the respective search range in the last column. For GS-Lin2, temperature thresholds (T_w, T_c) are higher for the C class (12.0-22.0°C), comparatively to D (3.5-9.5°C), following on the notion that leaf flushing is controlled by higher temperatures in warmer climates and by lower temperatures in colder ones. For the C climate, only warm spells of only 3 days would be required to onset GS, while more sustained cold spells of 7-9 days would be required to trigger EGS. On the other hand, the spell duration in D climates shows more variability. In particular, in warm summer and in severe winters subarctic continental climates (D*a and D*d) sustained warm and cold spells (11 and 14 days) are required to trigger GS transitions, while the other continental climates show more variability in the spell duration.

For GS-BC2, heat and cold cumulation ($t_{1,SGS}, t_{1,EGS}$) triggers late in most instances, with $t_1 < 30$ only for the SGS in D*b and D*c, and for the EGS in C*a. This is accompanied by relatively low critical thresholds (F_{crit}), indicating that heat and cold cumulation might last only a few days in the warmer and colder climates, respectively. The only climate with a conspicuous critical threshold is D*d, with 1014.5°C d required for SGS, cumulated from DOY 33. The temperature threshold for SGS ($T_{BF,SGS}$) is close to zero in all KG classes: in fact, a range including negative values could have led to lower parameters in many instances, but we preferred to limit parameter inference on biologically relevant values. Instead, for $T_{BF,EGS}$ there is no obvious pattern across KG classes and cold cumulation can start for relatively high temperatures (below 17.1-28.6°C, depending on the KG class). The photoperiod greatly influences heat cumulation in most instances, with the only exception of D*c, where a_{SGS} is close to zero (i.e., temperature cumulation is unaffected by photoperiod). On the other hand, the EGS seems to be controlled by low temperature in most climates rather than by the photoperiod, which appears to be relevant only in hot summer continental climates (D*a).

For GS-P, the probabilities predicted from the binary classifier show very clear profiles when plotted against T_{14d} and T_{60d} (Fig. S3a,d), indicating that absolute temperatures are the main determinant of GS likelihood. Temperature gradients (∇T_{14d}, Fig. S3b) and short-term variability (σ_{14d}, Fig. S3c) show weaker relationships. Nonetheless, these features contribute to transition identifications, as non-GS states can be associated with larger temperature slopes and with unstable temperatures. In general, continental climates (D) show much more defined GS probabilities compared to temperate climates (C), with absolute temperatures in the D classes that allow to identify sharp GS transitions, whereas relationships in the C classes are noticeably noisier. The GS-P hyperparameters associated with the probability thresholds were estimated separately for each KG class (

https://doi.org/10.5194/egusphere-2026-1710
Preprint. Discussion started: 8 April 2026

Table 1). The mean probability threshold μ is generally higher for SGS (0.537-0.606) than for EGS (0.373-0.554). As the chances to shift from a GS state to a non-GS state are approximately given by the complementary threshold probability for EGS ($1 - \mu_{EGS}$), this indicates a symmetric behaviour in temperature-probability relationships in signalling phenological transitions. However, the 95% CIs of μ (Table S2) suggest that SGS conditions are more tightly constrained by temperature (CI range: 0.013–0.088) compared to EGS (0.025–0.117), indicating a smaller uncertainty in predicting transitions into GS than out of it. The between-node variances τ^2 show larger values for SGS (0.049-0.058) than for EGS (0.025-0.39), suggesting that the degree of spatial heterogeneity in transition behaviour is larger for SGS than for EGS. Within-node variances σ^2 are lower in temperate (0.013-0.022) than continental climates (0.032-0.051). This likely reflects the sharper temperature-probability profiles over D climates, where even small temperature differences correspond to more diverse sampled GS probabilities compared to C climates. This is also coherent with the width of the 95% CIs, which indicate that the uncertainty of this parameter is consistently lower in C climates (0.004-0.007) than in D climates (0.006-0.023). Similar features are found in the node-level probability thresholds p_{SGS} and p_{EGS} (Figure S4).

GS-Lat regression parameters behave as expected in the NH. In fact, here, for each degree north, SGS delays by 1.93 d and EGS anticipates by 1.14 d; for each km in altitude SGS delays by 10.07 d and EGS anticipates by 9.94 d. However, for the SH, this pattern does not fully hold. While for EGS the expected sign is found (-1.28 d/°S, and -5.85 d/km), the SGS parameters have counterintuitive signs. This is likely due to the topographical features of South America, where the Andes span a wide latitudinal range and distort the expected relationship by concentrating high-elevation forests closer to the equator. The disrupted relationship between phenological transitions, latitude and elevation is also implied by the wider 95% CIs in the SH, especially for the coefficient related to elevation (Table S2; SGS: [-8.97,-7.94]; EGS: [-6.29,-5.42]).

https://doi.org/10.5194/egusphere-2026-1710
Preprint. Discussion started: 8 April 2026

Table 1. The parameter values for the GS-Lin2, GS-BC2, and GS-P models for each KG classification, and regression coefficients for GS-Lat for each hemisphere. The meaning of symbols can be found in the model description section (2.2).

	C*a	C*b ∨ C*c	D*a	D*b	D*c	D*d	Search Range
GS-Lin2							
T_w	22.0	13.9	5.9	7.7	6.5	3.5	[0, 30]
n_w	3	3	11	5	7	14	[3, 15]
T_c	18.8	12.0	4.9	6.5	7.2	9.5	[0, 30]
n_c	9	7	4	13	7	11	[3, 15]
GS-BC2							
$t_{1,SGS}$	88	64	63	1	11	33	[1,100]
$F_{crit,SGS}$	110.3	100.0	437.7	437.3	218.3	1014.5	[100, 5000]
$T_{BF,SGS}$	0.0	0.1	0.4	0.0	0.1	0.5	[0, 30]
a_{SGS}	4.9	2.6	3.8	4.2	0.2	2.8	[0, 5]
$t_{1,EGS}$	17	88	77	49	59	49	[1, 100]
$F_{crit,EGS}$	-558.8	-694.1	-568.5	-461.6	-100.6	-198.8	[-5000, -100]
$T_{BF,EGS}$	24.6	28.6	25.7	18.3	17.1	22.9	[0, 30]
a_{EGS}	-0.1	-0.2	-1.2	0.0	0.0	0.0	[-5, 0]
GS-P							
μ_{SGS}	0.555	0.537	0.608	0.585	0.583	0.606	
τ^2_{SGS}	0.503	0.554	0.373	0.406	0.488	0.458	
σ^2_{SGS}	0.056	0.054	0.049	0.052	0.058	0.045	
μ_{EGS}	0.039	0.038	0.031	0.028	0.038	0.025	
τ^2_{EGS}	0.013	0.017	0.033	0.032	0.034	0.040	
σ^2_{EGS}	0.018	0.022	0.051	0.050	0.045	0.051	
GS-Lat	**NH**			**SH**			
	Estimate	SE	t-stat	Estimate	SE	t-stat	
$\beta_{0,SGS}$	19.41	0.11	172.48	297.89	0.62	479.31	
$\beta_{1,SGS}$	1.93	0.00	975.56	-1.38	0.02	-78.47	
$\beta_{2,SGS}$	10.07	0.04	272.84	-8.45	0.26	-31.93	
$\beta_{0,EGS}$	342.14	0.17	2054.64	197.27	0.52	377.98	
$\beta_{1,EGS}$	-1.14	0.00	-389.01	-1.28	0.01	-86.78	
$\beta_{2,EGS}$	-9.94	0.05	-182.34	-5.85	0.22	-26.31	

https://doi.org/10.5194/egusphere-2026-1710
Preprint. Discussion started: 8 April 2026

3.2 GS models performance

Table 2 shows the SS for MAE and ρ of the three temperature-based models, indicating the percentage of improvements over the performance of the benchmarking model GS-Lat. SS are calculated over different KG classes and globally, using the performance metrics penalized over the number of failed instances (Table S3).

Table 2. The MAE and ρ Skill Scores (SS, [%]) over each KG class for the three models GS-Lin2, GS-BC2 and GS-P, computed from MAE and ρ penalized on the number of failed instances (Table S3).

Metric	KG classification	C*a	C*b ∨ C*c	D*a	D*b	D*c	D*d	Global
SS(MAE)				**SGS**				
	GS-Lin2	-90.1	-71.0	-38.6	-0.8	24.9	-9.8	-15.0
	GS-BC2	38.1	55.5	36.2	53.2	33.3	45.9	44.2
	GS-P	20.3	26.3	48.1	68.1	62.5	49.5	45.8
				EGS				
	GS-Lin2	-13.9	14.3	30.2	35.0	24.9	60.1	17.5
	GS-BC2	23.3	53.8	63.7	57.5	57.2	83.4	52.0
	GS-P	27.0	25.5	58.5	59.3	62.2	77.5	45.3
SS(ρ)				**SGS**				
	GS-Lin2	-349.3	-297.6	45.2	35.0	-66.6	22.7	-106.1
	GS-BC2	16.7	26.9	58.3	56.9	-8.2	31.8	44.4
	GS-P	-33.2	-131.5	78.6	79.9	63.0	56.8	8.5
				EGS				
	GS-Lin2	-25.5	-83.9	-1.6	28.4	18.3	-45.3	-4.8
	GS-BC2	-5.1	57.8	23.0	49.2	36.5	-4.7	49.3
	GS-P	17.4	-5.1	32.8	59.2	66.4	-6.3	29.8

At a global scale, GS-Lat shows MAE values of 20.6 d for SGS and of 25.2 d for EGS. GS-P leads to large improvements globally (SGS: 45.8%, EGS: 45.3%), corresponding to MAE of 11.2 d for SGS and of 13.8 d for EGS. On the other hand, GS-Lin2 shows smaller improvements only during EGS (17.5%), while GS-BC2 performs as good as GS-P in both transitions (SGS: 44.2%, EGS: 52.0%). In terms of linear correlation, GS-Lat shows fairly large values for SGS (0.84) and lower values for EGS (0.65). GS-P shows small improvements globally for SGS (8.5%), and larger for EGS (29.8%). On the other hand, GS-Lin2 shows considerable degradation in SGS, while GS-BC2 leads to large improvements (SGS: 44.4%, EGS: 49.3%).

Models perform substantially better over continental than over temperate climates. For MAE, GS-P shows improvements between 48.1 and 77.5% over D classes and between 20.3 and 27.0% over C classes. For this model, MAE values can be considered negligible for most applications over continental climates (3.0-7.6 days), whereas values are larger over temperate climates (22.3-39.2 days). GS-P also maintains high correlation across all KG classes, with the only noticeable exception in the EGS for D*d (0.32), and with improved or comparable performances to GS-Lat. Overall, GS-Lin2 performs as well as GS-Lat, with some improvements over D climates in terms of MAE, but noticeable degradations in the SGS over

https://doi.org/10.5194/egusphere-2026-1710
Preprint. Discussion started: 8 April 2026

C climates for both MAE and ρ. GS-BC2 performs similarly to GS-P across KG classes, outperforming it in temperate climates, while GS-P remains superior in continental regions.

In terms of systematic biases, GS-P displays essentially no bias over continental climates, with a MB between -1.4 and 1.4 d. On the other hand, over C climates, there is a consistent early bias of 15.4-18.2 d for SGS and a late bias of 17.3-21.7 d for EGS. GS-Lin2 shows a late bias in the SGS across all KG-classes, with a minimum 0.5 d over D*d and a maximum 78.7 d over C*b and C*c. On the other hand, the MB range is smaller for EGS, ranging between -7.1 and 6.3 d. GS-BC2 shows virtually no bias across all KG classes and transitions, with the only exception of C*a (13.9 d). As explained in section 2.3, temperature-based models may fail to produce valid predictions under certain conditions. GS-P provides robust predictions across all KG classes, whereas GS-Lin2 fails to produce valid estimates for 31.78% and 31.73% of cases for SGS and EGS, respectively, over C*b and C*c. GS-BC2 also shows limitations, failing to generate 16.95% of predictions over C*a.

The evaluation described above is reflected in the average MB maps (Figure 1), with GS-Lin2 showing the largest magnitudes, together with GS-Lat, and followed by GS-P and GS-BC2. All GS models show similar critical areas: larger biases, either early or late, are generally observed over parts of Central America, and Southeastern China, South and South-East Asia. Beyond these areas, GS-Lin2 and GS-Lat generally shows larger biases than GS-BC2 and GS-P, especially across the SH, and over parts of the US and Europe. Figure S5 shows the mean across the years of the error spread p_{95} (section 2.4) for each of the GS models. This figure further confirms the spatial features of the MB. For GS-P (Fig. S5e,f) the uncertainty in the prediction is small for a large part of the NH, while the largest error spread is observed over India and in South America, especially in the equatorial Andes. GS-BC2 performs similarly to GS-P, although the uncertainty appears to be generally greater for SGS, and smaller across the SH. The remaining two models show similar spatial patterns, although with varying magnitude.

A final component of model evaluation should consider the predictive capabilities of GS models outside their ERA5 calibration range. Figure S6 shows the percentage of failed predictions over C and D climates using UKESM1-0-LL historical (1850-2014) and SSP5-8.5 (2015-2100) temperatures. UKESM1-0-LL is known to have a substantial cold bias over the historical period (Sellar et al. 2019) and to project one of the strongest warming trends among CMIP6 ESMs (Zelinka et al. 2020). As such, this ESM constitutes a suitable testing ground for the robustness of the GS models. Across all models, failed predictions are more frequent during the historical period than under SSP5-8.5, likely because temperatures colder than the training range impair model performances. Under global warming, UKESM1-0-LL temperatures become closer to ERA5 conditions, reducing the number of failed predictions. Among GS models, GS-BC2 has the highest failure rate across all the 250 years, with about 15-25% failed instances in the historical period, and 5-15% of failed instances under SSP5-8.5, whereas GS-P and GS-Lin2 failure ratios are much lower (5-10% over the historical period, <2.5% under SSP5-8.5).

https://doi.org/10.5194/egusphere-2026-1710
Preprint. Discussion started: 8 April 2026

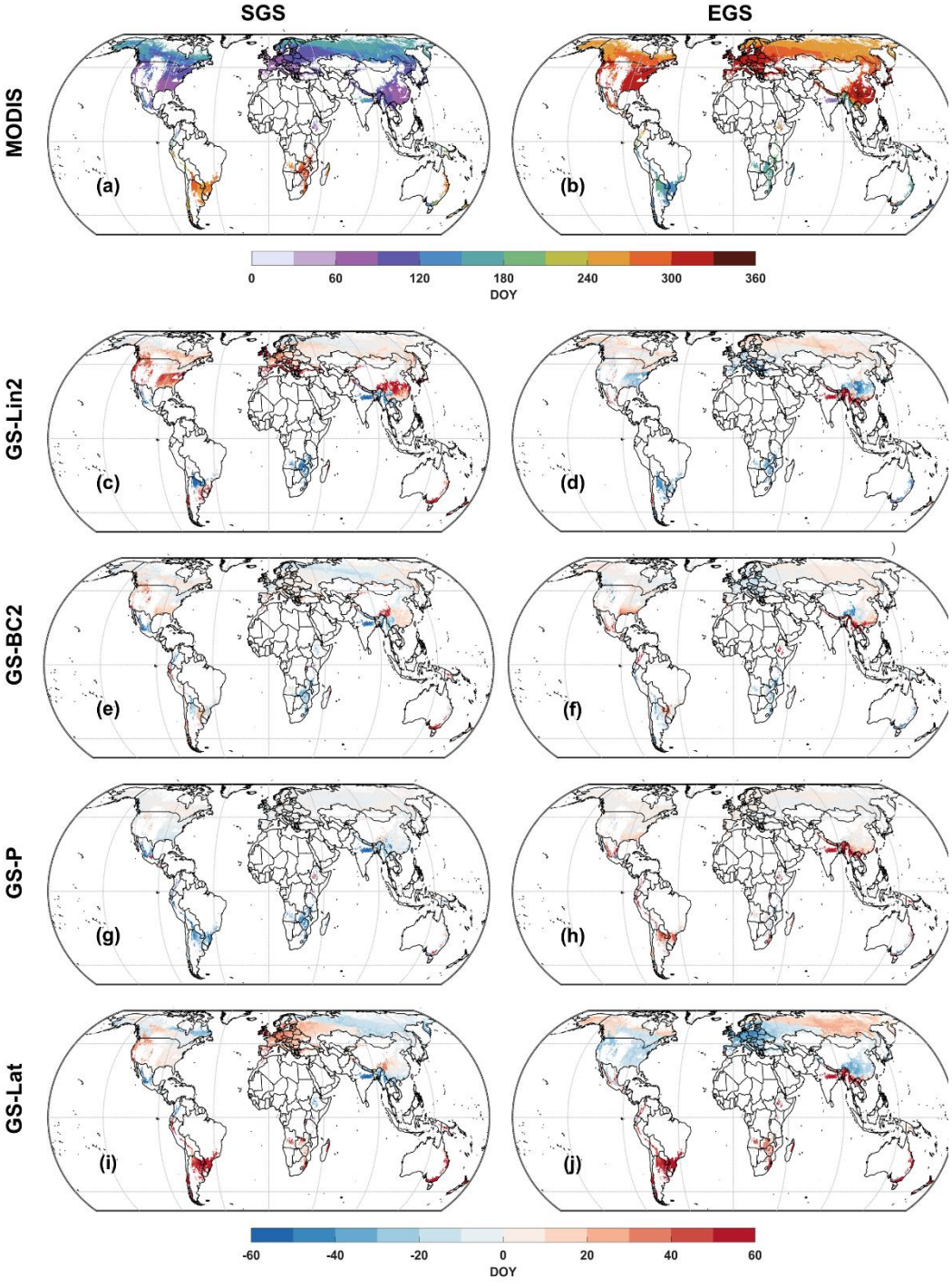

Figure 1. MODIS SGS and EGS maps (a,b) and MB maps (Model – MODIS) for SGS (first column) and EGS (second column) as estimated by GS-Lin2 (c,d), GS-BC2 (e,f), GS-P (g,h) and GS-Lat (i,j).

https://doi.org/10.5194/egusphere-2026-1710
Preprint. Discussion started: 8 April 2026

3.3 Forest GS from 1850 to 2100 using the GS-P model

Figure 2 shows the SGS and EGS calculated by GS-P using UKESM1-0-LL near-surface air temperatures, for the preindustrial period (1850-1879), and as end-of-century mean differences relative to the preindustrial conditions under SSP1-2.6 and SSP5-8.5. The SGS and EGS are computed only over grid nodes with at least 10% forest cover in at least one year of any experiment.

During the preindustrial period (historical experiment, 1850-1879; Figure 2Figure 3a,b), the SGS and EGS show the expected large-scale behaviour across most of the globe: for a given year, SGS precedes EGS in the NH, while the opposite ordering is observed in the SH, and transitions timings broadly follow the latitude-elevation gradient. Only limited regions, such as north-western Australia and the tropical Andes, exhibit an inversion of the typical SGS-EGS order. Under SSP1-2.6 and SSP5-8.5, global warming leads to earlier SGS and later EGS (Figure 2c-f). In this regard, Table S4 shows the SGS and EGS values over different KG classes, as identified in the preindustrial period (1850-1879), and relative changes over present-day (1985-2014) and end 21st century periods (2071-2100). In the NH, from the preindustrial to the present-day period, on average, the SGS is estimated to anticipate from 0.5 to 3.1 d and the EGS to delay from 0.5 to 3.4 d, depending on the KG class considered (Table S4). Consequently, the GS duration is estimated to increase by 1.4 to 5.4 d on average, with the preindustrial C-class areas experiencing the larger effect. For future periods, these shifts become consistently more evident across most classes and scenarios, with the largest differences over preindustrial D-class areas: between 2071 and 2100, the GS is projected to last up to 48.5 d longer on average (compared to the preindustrial period) under SSP1-2.6, and up to 85.4 d longer under SSP5-8.5. In the SH, similar trends occur over preindustrial C*a and D*c climates, with the GS becoming on average 10.2-13.2 d longer at present day, and 34.8-53.9 d longer at the end of the century under SSP5-8.5. However, for C*b and C*c climates, a slight shortening of the GS is projected for all future scenarios.

Beyond SGS and EGS shifts, two additional features emerge. First, the regions near the boundaries of KG classes undergo climatic shifts, leading to changes in the model parameterization (e.g., a transition from a D to a C GS regime) that mitigate the end-of-century differences in SGS and EGS relative to preindustrial conditions; Second, GS-P identifies *GS-flipping* regions, where the SGS-EGS ordering is inverted compared to the preindustrial period, causing observed large differences. This exclusively occurs at the interface between temperate (C) and warmer (A or B) climates. For instance, in the southeastern United States, GS-P simulates the SGS and EGS occurring in February-March and November-December during the preindustrial period, respectively; at the end of the 21st century, under SSP5-8.5 (Figure 2e), SGS is estimated to occur in July-August, while EGS in May-June, following a regime that is typically observed in the SH or in monsoon-controlled regions, such as South and Southeast Asia (Figure 1a,b). At the preindustrial period, prediction uncertainty is remarkably low (<10 d) for most of the NH but increases in several SH regions (Fig. S5). In addition, for the end-of-century period, the largest error spread occurs in areas associated with the GS-flipping behaviour, reflecting predictions that strongly deviate from the calibration range of SGS and EGS.

https://doi.org/10.5194/egusphere-2026-1710
Preprint. Discussion started: 8 April 2026

Figure 3 shows the yearly averages of SGS and EGS calculated in the two hemispheres by GS-P, from 1850 to 2100, under SSP1-2.6 and SSP5-8.5. GS-flipping nodes were excluded from the calculation in order to produce consistent averages for the whole 1850-2100 period. Stable average SGS and EGS dates are observed for both hemispheres from 1850 to around 1970. During the long and stable historical period (1850-1969), SGS and EGS average at around DOY 134.8 (15 May) and 272.5 (29 September) in the NH, and around DOY 244.6 (2 September) and 156.8 (6 June) in the SH, respectively. Changes in phenological transitions timing appear to emerge from 1970s. Table 3 shows shift ratios in yearly mean SGS and EGS timing calculated over the late-historical period (1970-2014) and over the end of the 21st century (2071-2100), across different areas as identified by preindustrial KG classes. During the 1970-2014 period, SGS is estimated to anticipate at a statistically significant ratio between 0.7 and 3.5 d decade^{-1}, and EGS to delay between 1.4 and 3.6 d decade^{-1} over those areas with C, D*c and D*d climates during preindustrial period. Projection for the 2071-2100 period under SSP1-2.6, shows that these ratios may be dampened, enhanced or no longer significant, depending on the region. In contrast, these effects are usually greatly enhanced for the same period under SSP5-8.5, with ratios reaching values of -9.1 d decade^{-1} for SGS and +9.2 d decade^{-1} for EGS, over the preindustrial D*d area. Beyond these regions, preindustrial D*a areas typically display the opposite behaviour, with delayed SGS and anticipated EGS, while ratios over preindustrial D*b areas are either not significant or small comparatively to the other regions.

Figure S7-S8 shows the yearly averages calculated via GS-Lin2 and GS-BC2. For the NH, they both show a behaviour similar to GS-P, although they differ in absolute timing of transitions and magnitude of transition shifts. For the SH, it should be noticed that GS-Lin2 produces ill-defined results, with an extremely large prediction interval for both SGS and EGS.

https://doi.org/10.5194/egusphere-2026-1710
Preprint. Discussion started: 8 April 2026

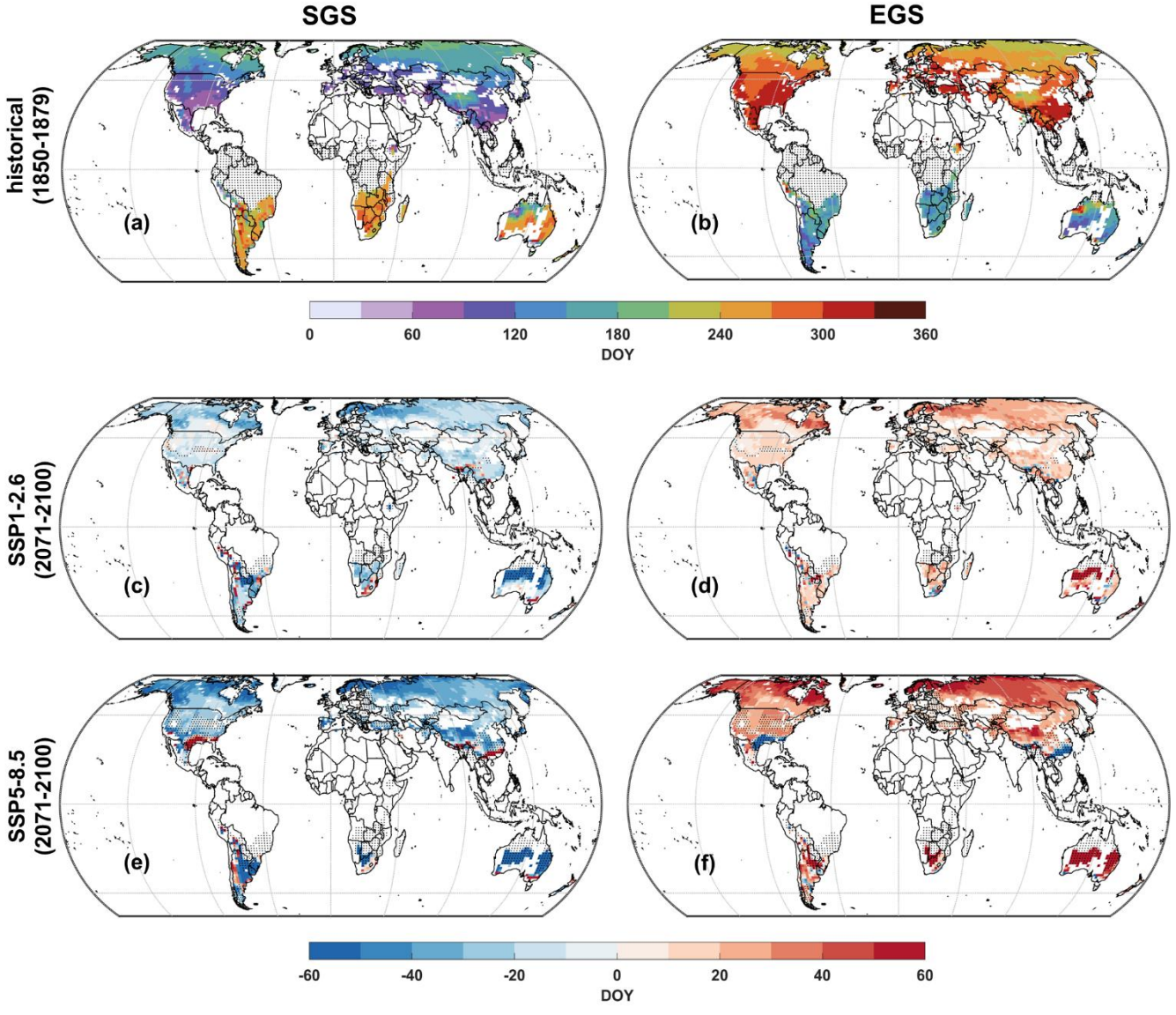

Figure 2. Mean SGS (first column) and EGS (second column) calculated by GS-P for the preindustrial period 1850-1879 (a,b), and differences between the preindustrial period mean and the 2071-2100 mean, under SSP1-2.6 (c,d) and under SSP5-8.5 (e,f). In (a,b) the dots indicate nodes with at least one year of the KG class A. In (c-f) dots indicate nodes that transitioned between main classes (i.e. from C and D to A, B, or C) or that present *GS-flipping* behaviour, i.e. those locations where the order of SGS and EGS switches, comparatively to the preindustrial period. SGS and EGS are calculated only over the nodes with at least 10% primary or secondary forest cover in at least one year of any experiment.

https://doi.org/10.5194/egusphere-2026-1710
Preprint. Discussion started: 8 April 2026

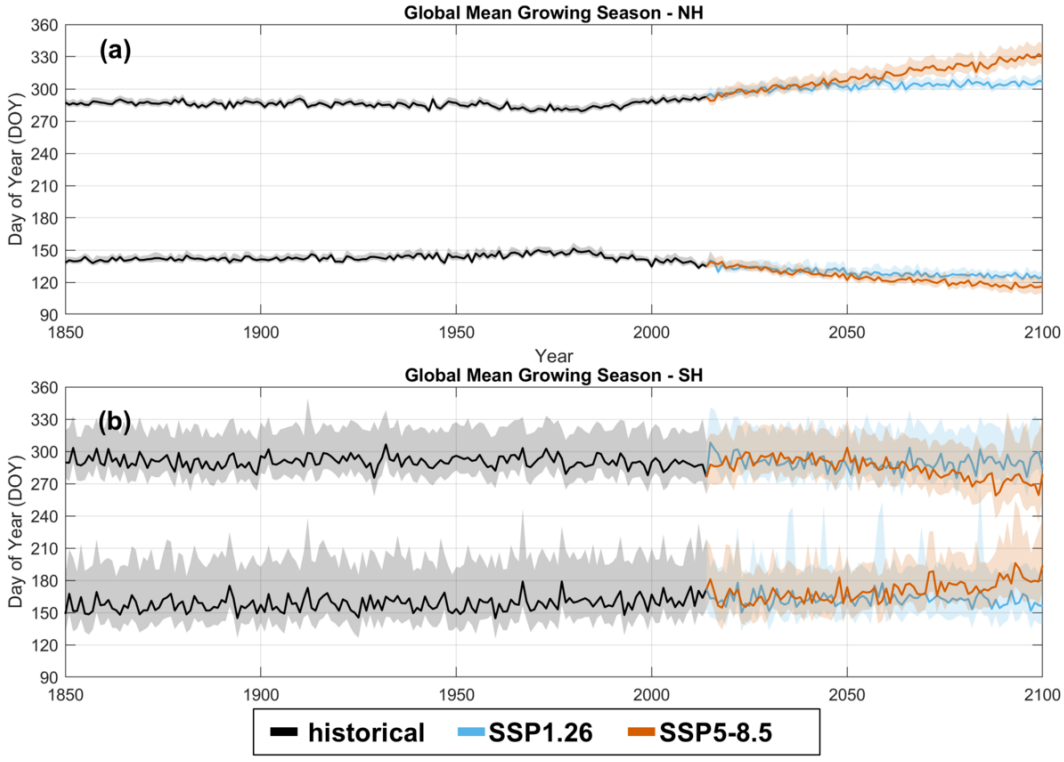

Figure 3. Area-weighed yearly averages of SGS and EGS from 1850 to 2100, for the Northern (a) and the Southern Hemisphere (c), as simulated by GS-P over UKESM1-0-LL temperatures, only considering non-KG transition, non-flipping nodes. In the NH, the bottom lines refer to the SGS, while the top lines refer to the EGS. In the SH, the order is reversed. Shaded areas indicate the 95% confidence interval.

Table 3. Ratios of changes ([d decade^{-1}]) in yearly mean SGS and EGS over different KG classes during the historical years when shifts in phenological transitions started to emerge (1970-2014), and during the end of the 21st century (2071-2100) under different scenarios. Here, KG classes identify nodes predominantly within that class during the preindustrial period (1850-1879). Starred numbers indicate p<0.05 for testing non-null mean ratios.

	Period	Experiment	C*a	C*b ∨ C*c	D*a	D*b	D*c	D*d	C	D	Global
SGS	1970-2014	historical	-3.2*	-3.3*	1.0*	0.0	2.8*	10.4*	-3.2*	-1.5*	-2.1*
	2071-2100	SSP1-2.6	-2.2*	-0.1	0.8*	0.1	-0.7*	-1.6*	-1.8*	-1.1*	-1.2*
	2071-2100	SSP3-7.0	-5.4*	-2.4*	0.6*	-0.2	-2.3*	-2.2	-5.4*	-2.6*	-3.1*
	2071-2100	SSP5-8.5	-6.1*	-5.4*	0.3*	-0.1	-3.5*	-3.5	-6.9*	-3.2*	-3.9*
EGS	1970-2014	historical	3.6*	2.3*	-0.7*	0.6*	1.3*	2.5*	3.4*	2*	2.4*
	2071-2100	SSP1-2.6	1.7*	-1.1*	-0.5*	0.1	1.0*	3.1*	1.4*	1.2*	1.1*
	2071-2100	SSP3-7.0	4.5*	-1.0*	0.0	0.5*	2.8*	6.5*	5.1*	3.1*	3.2*
	2071-2100	SSP5-8.5	4.7*	1.3*	0.4*	0.2	3.8*	8.3*	6.4*	3.8*	3.9*

https://doi.org/10.5194/egusphere-2026-1710
Preprint. Discussion started: 8 April 2026

4. Discussion

4.1 Historical trends and future projections of GS transitions

The GS-P model estimates of this study indicate that the SGS and EGS remained stable from 1850 until the 1970s. Following this period, systematic phenological shifts emerged: over 1970-2014, the average global rate of change was approximately -2.1 d decade^{-1} for the SGS and +2.4 d decade^{-1} for the EGS, with stronger shifts at higher latitudes compared to lower ones. Richardson et al. (2013) reviewed shifts of a similar order, although more extreme rates have been observed in specific biomes and time periods (e.g., up to +14 d decade^{-1} for boreal forests in recent decades). Consistent satellite-based estimates were also reported by (Park et al. 2016; Barichivich et al. 2013) who found a GS duration increase of +2.6 and +2.2 d decade^{-1}, respectively, in the last 4 decades. Most phenological trend studies focus on the NH, with far fewer addressing vegetation phenological phases in the SH. Moreover, studies in the SH typically focus on flowering dates rather than on the timing of SGS and EGS. From the few long-term studies performed for the SH, evidence of responses to global warming seems weaker and uncertain. For instance, while the review by Chambers et al. (2013) reported an average advance in flowering time of -5.6 d decade^{-1}, it should be also noted that 72% of the considered datasets lacked statistically significant trends. Similarly, Everingham et al. (2023) investigated species-specific responses in the shifts of flowering dates of 37 herbaceous taxa and found significant shifts in only 12 species.

For future periods, GS duration is projected to increase substantially across the NH, comparatively to the present-day climatological mean (1985-2014). By the end of the 21st century, GS duration is expected to increase by approximately 24.0-48.8 d over regions with preindustrial temperate climates and by 32.4-67.0 d over regions with preindustrial continental climates, depending on the scenario. In the SH, the response is generally weaker: GS duration increases by approximately 14.8-18.2 days in temperate climates and 26.7 days in continental climates under SSP5-8.5, while a shortening of 13.1 days is projected under SSP1-2.6 for the same continental regions. Although the shifts simulated by GS-P model may seem large, they are consistent with the physiological response to temperature reported in the literature. For instance, Liu et al. (2021) derived a GS length sensitivity to temperature of 8.22 d °C^{-1} (95% CI: [2.04, 14.40]) from controlled experiments data. Since CMIP6 projections indicate a global mean temperature increase of 1.18 °C and 7.20 °C century^{-1} under SSP1-2.6 and SSP5-8.5, respectively (Fan et al. 2020), this yields a mean GS lengthening of approximately 2.4–17.0 d century^{-1} for the low emissions scenario and 14.7–103.7 d century^{-1} for the high-emissions scenario.

Few studies have explicitly modelled future changes in SGS and EGS. Sun et al. (2025), using a boosted random forest (XGBoost) with 18 temperature-related predictors, estimated SGS timing up to 2100 for boreal vegetation in the NH (which broadly corresponds to D*c and D*d climate class in our study). They found much smaller shifts than those obtained in this study; from the beginning to the end of the 21st century, they estimated an advance of only 0.5–1.2 and 6.8-8.3 days under SSP1-2.6 and SSP5-8.5 respectively, depending on forest ecosystem type. In contrast, our estimates indicate earlier SGS shifts of 19.3–23.8 days under SSP1-2.6 and 37.3–41.7 days under SSP5-8.5. This discrepancy likely reflects differences in modelling assumptions. Sun et al. (2025) trained their model on present-day boreal ecosystems and extrapolated the resulting

https://doi.org/10.5194/egusphere-2026-1710
Preprint. Discussion started: 8 April 2026

eco-climatic relationships into the future, implicitly assuming the persistency of the ecosystem type. Consequently, their approach may underestimate future phenological shifts by applying boreal forest features in regions which are likely to experience transitions in vegetation types. In contrast, our approach constrains the ecosystem-climate relationship within KG climate classes, effectively assuming equilibrium between climate and ecosystem compositions, and allowing ecosystem behaviour to track projected climate change. However, because of the nearly simultaneous response to the evolving temperature regime, our approach may overestimate the rate or magnitude of future phenological regime shifts where changes in vegetation composition and physiological acclimation are slower than climate change.

4.2 Performance comparison between GS-P and other temperature-based models

Benchmarking the three temperature-based models against GS-Lat for 2001-2023 period allows for a broad ranking of their performance. The GS-Lin2 model, arguably the simplest among the three considered, performs on-par with GS-Lat in terms of prediction errors. GS-P and GS-BC2 show similar overall performance, although their skill varies across regions and KG climate classes. The most extensive spatial difference occurs in the SH, where GS-P shows more systematic bias than GS-BC2. However, a critical limitation of GS-BC2 is its substantial higher number of failed predictions compared to GS-P. This indicates poorer extrapolation capabilities for GS-BC2 and, overall, greater robustness of GS-P.

All temperature-based models show limited skills in reproducing GS transitions over the SH, suggesting that temperature alone may not provide sufficient explanatory power in these regions and that additional drivers of SGS and EGS timing may be more important. However, GS-P better represents climate-associated shifts in phenological transitions compared to other models. Since earlier SGS and later EGS are generally expected in a warmer climate, GS-P reasonably reproduces this behaviour in both hemispheres, while GS-Lin2 and GS-BC2 show some shortcomings (Figures S8, S9). A further advantage of GS-P is that computing CIs on the parameters is easy, whereas this task is not trivial for GS-Lin2 and GS-BC2 and requires computationally expensive workarounds. This allows for a more grounded interpretation of GS-P results, whereas any interpretation on the parameters for GS-BC2 and GS-Lin2 is mostly speculative.

GS-P can be readily compared with the similar model proposed by Sun et al. (2025): their boosted random forest model was evaluated to have an MAE smaller than 6 days in 41% of the nodes (across all vegetation types), whereas for GS-P this threshold is met by 88% of the nodes over the forested boreal region. GS-P can also be compared with other more traditional phenological models. Y. Mo et al. (2023) compared the predictive ability of 17 different phenological models in predicting the SGS in a variety of vegetation types along a north-south transect of Northeast Asia. In their study, model performance over mixed and deciduous needleleaf forests (the only two forest types considered) yielded RMSE values between 4.9 and 8.2 d and a correlation coefficient between 0.54 and 0.76. Over the same sub-domain, and evaluated over the 2001-2023 period, GS-P achieves an RMSE of 5.4 d and a correlation coefficient of 0.97, suggesting that our model performs among the top tier of traditional phenological models.

https://doi.org/10.5194/egusphere-2026-1710
Preprint. Discussion started: 8 April 2026

4.3 Ecological interpretation of GS-P parameters

GS-P parameters can be interpreted from an ecological perspective to understand the physiological features captured by the model. First, GS-P requires approximately symmetric temperature evidence to signal transitions into and out of the GS, as shown by the mean probability thresholds (μ) associated with each KG class, which represent the average level of temperature-based evidence required to trigger a phenological transition. However, the lower uncertainty associated with SGS thresholds indicates a stronger temperature-phenology signal in spring than in autumn. This finding aligns with previous literature, which consistently reported well-defined responses in spring phenology to temperatures, while autumn phenology often showed more variable and less predictable responses (Menzel et al. 2006; Richardson et al. 2013; Chmielewski and Rötzer 2001). Second, GS-P reveals that conditions for the SGS are more spatially heterogeneous than EGS conditions (between-node variance, τ^2). This suggests that spring phenology timing is a more critical period, potentially more sensitive to microclimatic conditions, species compositions and local adverse events such as late frosts. Third, GS-P shows that phenological transitions are more easily detected in continental (D) than in temperate (C) climates using only temperature-based criteria. This is suggested by the sharper temperature-probability profiles in the D climates than in the C climates and, consequently, by the larger values of the within-node variance (σ^2) of the probability thresholds: over continental regions, relatively small temperature changes lead to large changes in the probability of GS presence, consistently with stronger seasonality and less interference by other environmental factors. Instead, in temperate climates, noisy probability curves and lower predictive skill indicate that temperature provides a weaker signal for GS transitions in these regions. Here, σ^2 is lower than in D climates because the temperature-phenology structure is not as strong as in C climates, limiting the model's ability to resolve meaningful within-class differences.

Finally, the GS-flipping feature (section 3.3) identified at the interface between temperate and warmer climates requires careful interpretation. Rather than an actual ecological phenomenon (which is clearly unobserved in the real world), it should be interpreted as a model diagnostic signal of large-scale climate-ecosystem disequilibrium, indicating areas where transitional states anticipate a full shift to tropical or arid ecosystems, depending on the region's precipitation regime. Alternatively, the GS-flipping feature may just indicate that temperature-phenology relationships in these areas simply cannot be fully interpreted within the current modelling framework. Under the first hypothesis, these regions should be considered of particular interest for ecosystem conservation strategies, as rapid climate changes might outpace ecological adjustments, increasing the risk of biome reorganization or loss of ecosystem function (Seidl and Turner 2022). Instead, under the alternative hypothesis, they highlight a challenging scientific problem in phenology modelling and provide a valuable testing ground for improving our understanding of vegetation–climate interaction.

5 Conclusions

In this study, we developed and calibrated GS-P, a probabilistic, temperature-based model designed to estimate SGS and EGS of generic forest ecosystems at a global scale from 1850 to 2100. Our results indicate a general prolongation of the GS

https://doi.org/10.5194/egusphere-2026-1710
Preprint. Discussion started: 8 April 2026

under climate warming from the preindustrial period to the end of the 21st century, with an increase of approximately one month under SSP1-2.6 and two months under SSP5-8.5.

GS-P successfully identified key biological features of forest ecosystems, suggesting meaningful differences in phenological strategies. Specifically, SGS transitions are characterized by a more certain and spatially heterogeneous temperature signal, whereas autumn transitions are more homogeneous and exhibit greater uncertainty in their temperature dependence. Furthermore, GS-P identified regions where temperature-driven predictions become less reliable or less interpretable, suggesting a possible mismatch between climatic forcing and ecosystem response. These findings possibly indicate regions where ecosystems may be exposed to rapidly changing climatic conditions, highlighting priority zones of ecological interest for future research, monitoring and conservation efforts.

The model was directly compared with two other temperature-based models from the literature, which were reformulated and recalibrated for the purposes of this study. While GS-Lin2 provided reasonable large-scale estimates in the NH, it exhibited high uncertainty in the SH, performing broadly similarly to coordinate-based models. GS-BC2 achieved comparable or slightly better accuracy than GS-P but showed very limited extrapolation capabilities, likely due to overparameterization, which restricts its utility mainly to geographically and climatically constrained regions and present-day conditions. Instead, GS-P proved suitable for large-scale and climate-change applications, although its assumption of synchronous climate-GS regimes transitions may overestimate future phenological shifts. Furthermore, GS-P also enabled the calculation of confidence intervals for model parameters, providing rigorous insight into phenological processes, whereas GS-BC2 and GS-Lin2 offer limited interpretability in this regard.

Finally, a comparison with existing models from previous research revealed that regarding predictive accuracy, ecological realism, and general applicability, GS-P ranks among the top-performing phenological models for forest growing season simulations.

Data availability

The GS-P output described in this study is publicly available at https://doi.org/10.5281/zenodo.19224585.

Author contribution

All authors: conceptualization, methodology, original draft preparation, review and editing; PRG: data curation, formal analysis, investigation, software, validation; RM: visualization; GG: funding acquisition, project administration; GG and RM: supervision.

https://doi.org/10.5194/egusphere-2026-1710
Preprint. Discussion started: 8 April 2026

Competing interests

The authors declare that they have no conflict of interest.

Financial support

This work was supported by the FUTUROZ Project of the national "5x1000 funding to research" (Ministry of University and Research), by Catholic University of the Sacred Heart in the frame of its Programs of promotion and dissemination of the scientific research [Funding line D3.1].

References

Barichivich, Jonathan, Keith R. Briffa, Ranga B. Myneni, et al. 2013. "Large-scale Variations in the Vegetation Growing Season and Annual Cycle of Atmospheric CO_2 at High Northern Latitudes from 1950 to 2011." *Global Change Biology* 19 (10): 3167–83. https://doi.org/10.1111/gcb.12283.

Blümel, Klaus, and Frank-M. Chmielewski. 2012. "Shortcomings of Classical Phenological Forcing Models and a Way to Overcome Them." *Agricultural and Forest Meteorology* 164 (October): 10–19. https://doi.org/10.1016/j.agrformet.2012.05.001.

Bonan, Gordon B. 2016. *Ecological Climatology: Concepts and Applications*. Third edition. Cambridge University Press.

Buendía, C., A. Kleidon, and A. Porporato. 2010. "The Role of Tectonic Uplift, Climate, and Vegetation in the Long-Term Terrestrial Phosphorous Cycle." *Biogeosciences* 7 (6): 2025–38. https://doi.org/10.5194/bg-7-2025-2010.

Chambers, Lynda E., Res Altwegg, Christophe Barbraud, et al. 2013. "Phenological Changes in the Southern Hemisphere." *PLoS ONE* 8 (10): e75514. https://doi.org/10.1371/journal.pone.0075514.

Chapin, F. Stuart, Pamela A. Matson, and Peter M. Vitousek. 2011. *Principles of Terrestrial Ecosystem Ecology*. Springer New York. https://doi.org/10.1007/978-1-4419-9504-9.

Chen, Tianqi, and Carlos Guestrin. 2016. "XGBoost: A Scalable Tree Boosting System." *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, August 13, 785–94. https://doi.org/10.1145/2939672.2939785.

Chmielewski, Frank-M., and Thomas Rötzer. 2001. "Response of Tree Phenology to Climate Change across Europe." *Agricultural and Forest Meteorology* 108 (2): 101–12. https://doi.org/10.1016/S0168-1923(01)00233-7.

Chuine, Isabelle, Iñaki Garcia De Cortazar-Atauri, Koen Kramer, and Heikki Hänninen. 2013. "Plant Development Models." In *Phenology: An Integrative Environmental Science*, edited by Mark D. Schwartz. Springer Netherlands. https://doi.org/10.1007/978-94-007-6925-0_15.

Claverie, Martin, Junchang Ju, Jeffrey G. Masek, et al. 2018. "The Harmonized Landsat and Sentinel-2 Surface Reflectance Data Set." *Remote Sensing of Environment* 219 (December): 145–61. https://doi.org/10.1016/j.rse.2018.09.002.

Delahaye, Daniel, Supatcha Chaimatanan, and Marcel Mongeau. 2019. "Simulated Annealing: From Basics to Applications." In *Handbook of Metaheuristics*, edited by Michel Gendreau and Jean-Yves Potvin, vol. 272. International Series in Operations Research & Management Science. Springer International Publishing. https://doi.org/10.1007/978-3-319-91086-4_1.

Emberson, Lisa. 2020. "Effects of Ozone on Agriculture, Forests and Grasslands." *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 378 (2183): 20190327. https://doi.org/10.1098/rsta.2019.0327.

Everingham, Susan E., Raymond A. J. Blick, Manon E. B. Sabot, Eve Slavich, and Angela T. Moles. 2023. "Southern Hemisphere Plants Show More Delays than Advances in Flowering Phenology." *Journal of Ecology* 111 (2): 380–90. https://doi.org/10.1111/1365-2745.13828.

https://doi.org/10.5194/egusphere-2026-1710
Preprint. Discussion started: 8 April 2026

Eyring, Veronika, Sandrine Bony, Gerald A. Meehl, et al. 2016. "Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) Experimental Design and Organization." *Geoscientific Model Development* 9 (5): 1937–58. https://doi.org/10.5194/gmd-9-1937-2016.

Fan, Xuewei, Qingyun Duan, Chenwei Shen, Yi Wu, and Chang Xing. 2020. "Global Surface Air Temperatures in CMIP6: Historical Performance and Future Changes." *Environmental Research Letters* 15 (10): 104056. https://doi.org/10.1088/1748-9326/abb051.

Fowler, David, Mhairi Coyle, Ute Skiba, et al. 2013. "The Global Nitrogen Cycle in the Twenty-First Century." *Philosophical Transactions of the Royal Society B: Biological Sciences* 368 (1621): 20130164. https://doi.org/10.1098/rstb.2013.0164.

Friedl, Mark A., Damien Sulla-Menashe, Bin Tan, et al. 2010. "MODIS Collection 5 Global Land Cover: Algorithm Refinements and Characterization of New Datasets." *Remote Sensing of Environment* 114 (1): 168–82. https://doi.org/10.1016/j.rse.2009.08.016.

Fujimori, Shinichiro, Tomoko Hasegawa, Toshihiko Masui, et al. 2017. "SSP3: AIM Implementation of Shared Socioeconomic Pathways." *Global Environmental Change* 42 (January): 268–83. https://doi.org/10.1016/j.gloenvcha.2016.06.009.

Garnot, Vivien Sainte Fare, Lynsay Spafford, Jelle Lever, et al. 2025. "Deep Learning Meets Tree Phenology Modelling: PHENOFORMER versus Process-based Models." *Methods in Ecology and Evolution* 16 (7): 1489–506. https://doi.org/10.1111/2041-210X.70037.

Gelman, Andrew, John B. Carlin, Hal S. Stern, David B. Dunson, Aki Vehtari, and Donald B. Rubin. 2025. *Bayesian Data Analysis Third Edition (with Errors FIxed as of 20 February 2025)*.

Good, Peter, Alistair Sellar, Yongming Tang, et al. 2019a. "MOHC UKESM1.0-LL Model Output Prepared for CMIP6 ScenarioMIP Ssp126." Version 20250528. With Peter Good, Matthew Mizielinski, Mark Elkington, et al. Earth System Grid Federation. Application/x-netcdf. https://doi.org/10.22033/ESGF/CMIP6.6333.

Good, Peter, Alistair Sellar, Yongming Tang, et al. 2019b. "MOHC UKESM1.0-LL Model Output Prepared for CMIP6 ScenarioMIP Ssp370." Version 20250527. With Peter Good, Matthew Mizielinski, Mark Elkington, et al. Earth System Grid Federation. Application/x-netcdf. https://doi.org/10.22033/ESGF/CMIP6.6347.

Good, Peter, Alistair Sellar, Yongming Tang, et al. 2019c. "MOHC UKESM1.0-LL Model Output Prepared for CMIP6 ScenarioMIP Ssp585." Version 20250528. With Peter Good, Matthew Mizielinski, Mark Elkington, et al. Earth System Grid Federation. Application/x-netcdf. https://doi.org/10.22033/ESGF/CMIP6.6405.

Hersbach, Hans, Bill Bell, Paul Berrisford, et al. 2020. "The ERA5 Global Reanalysis." *Quarterly Journal of the Royal Meteorological Society* 146 (730): 1999–2049. https://doi.org/10.1002/qj.3803.

Hufkens, Koen, David Basler, Tom Milliman, Eli K. Melaas, and Andrew D. Richardson. 2018. "An Integrated Phenology Modelling Framework in R." *Methods in Ecology and Evolution* 9 (5): 1276–85. https://doi.org/10.1111/2041-210X.12970.

Körner, Christian, Patrick Möhl, and Erika Hiltbrunner. 2023. "Four Ways to Define the Growing Season." *Ecology Letters* 26 (8): 1277–92. https://doi.org/10.1111/ele.14260.

Kriegler, Elmar, Nico Bauer, Alexander Popp, et al. 2017. "Fossil-Fueled Development (SSP5): An Energy and Resource Intensive Scenario for the 21st Century." *Global Environmental Change* 42 (January): 297–315. https://doi.org/10.1016/j.gloenvcha.2016.05.015.

Linderholm, Hans W., Alexander Walther, and Deliang Chen. 2008. "Twentieth-Century Trends in the Thermal Growing Season in the Greater Baltic Area." *Climatic Change* 87 (3–4): 405–19. https://doi.org/10.1007/s10584-007-9327-3.

Liu, Huiying, Chunyan Lu, Songdan Wang, Fei Ren, and Hao Wang. 2021. "Climate Warming Extends Growing Season but Not Reproductive Phase of Terrestrial Plants." *Global Ecology and Biogeography* 30 (5): 950–60. https://doi.org/10.1111/geb.13269.

LRTAP Convention. 2017. "Chapter III: Mapping Critical Level for Vegetation." In *Modelling and Mapping Manual*. https://icpvegetation.ceh.ac.uk.

Meier, Michael, and Christof Bigler. 2023. "Process-Oriented Models of Autumn Leaf Phenology: Ways to Sound Calibration and Implications of Uncertain Projections." *Geoscientific Model Development* 16 (23): 7171–201. https://doi.org/10.5194/gmd-16-7171-2023.

https://doi.org/10.5194/egusphere-2026-1710
Preprint. Discussion started: 8 April 2026

Melaas, Eli K., Mark A. Friedl, and Zhe Zhu. 2013. "Detecting Interannual Variation in Deciduous Broadleaf Forest Phenology Using Landsat TM/ETM+ Data." *Remote Sensing of Environment* 132 (May): 176–85. https://doi.org/10.1016/j.rse.2013.01.011.

Menzel, Annette, Tim H. Sparks, Nicole Estrella, et al. 2006. "European Phenological Response to Climate Change Matches the Warming Pattern." *Global Change Biology* 12 (10): 1969–76. https://doi.org/10.1111/j.1365-2486.2006.01193.x.

Miralles, Diego G., Jordi Vilà-Guerau De Arellano, Tim R. McVicar, and Miguel D. Mahecha. 2025. "Vegetation–Climate Feedbacks across Scales." *Annals of the New York Academy of Sciences* 1544 (1): 27–41. https://doi.org/10.1111/nyas.15286.

Mo, Dengkui, Enping Yan, Hui Lin, Hua Sun, Jiping Li, and Guozhen Zhang. 2012. "Development and Validation of 2-Band EVI with MODIS Data in Southeast China." *Proceedings of 2012 International Conference on Measurement, Information and Control*, May, 88–91. https://doi.org/10.1109/MIC.2012.6273306.

Mo, Yunhua, Jing Zhang, Hong Jiang, and Yongshuo H. Fu. 2023. "A Comparative Study of 17 Phenological Models to Predict the Start of the Growing Season." *Frontiers in Forests and Global Change* 5 (January): 1032066. https://doi.org/10.3389/ffgc.2022.1032066.

Murphy, Allan H. 1988. "Skill Scores Based on the Mean Square Error and Their Relationships to the Correlation Coefficient." *Monthly Weather Review* 116 (12): 2417–24. https://doi.org/10.1175/1520-0493(1988)116%253C2417:SSBOTM%253E2.0.CO;2.

Park, Taejin, Sangram Ganguly, Hans Tømmervik, et al. 2016. "Changes in Growing Season Duration and Productivity of Northern Vegetation Inferred from Long-Term Remote Sensing Data." *Environmental Research Letters* 11 (8): 084001. https://doi.org/10.1088/1748-9326/11/8/084001.

Peel, M. C., B. L. Finlayson, and T. A. McMahon. 2007. "Updated World Map of the K¨oppen-Geiger Climate Classification." *Hydrol. Earth Syst. Sci.*

Richardson, Andrew D., Trevor F. Keenan, Mirco Migliavacca, Youngryel Ryu, Oliver Sonnentag, and Michael Toomey. 2013. "Climate Change, Phenology, and Phenological Control of Vegetation Feedbacks to the Climate System." *Agricultural and Forest Meteorology* 169 (February): 156–73. https://doi.org/10.1016/j.agrformet.2012.09.012.

Seidl, Rupert, and Monica G. Turner. 2022. "Post-Disturbance Reorganization of Forest Ecosystems in a Changing World." *Proceedings of the National Academy of Sciences* 119 (28): e2202190119. https://doi.org/10.1073/pnas.2202190119.

Sellar, Alistair A., Colin G. Jones, Jane P. Mulcahy, et al. 2019. "UKESM1: Description and Evaluation of the U.K. Earth System Model." *Journal of Advances in Modeling Earth Systems* 11 (12): 4513–58. https://doi.org/10.1029/2019MS001739.

Simpson, D., A. Benedictow, H. Berge, et al. 2012. "The EMEP MSC-W Chemical Transport Model – Technical Description." *Atmospheric Chemistry and Physics* 12 (16): 7825–65. https://doi.org/10.5194/acp-12-7825-2012.

Stuart Chapin III, F., Jack McFarland, A. David McGuire, Eugenie S. Euskirchen, Roger W. Ruess, and Knut Kielland. 2009. "The Changing Global Carbon Cycle: Linking Plant–Soil Carbon Dynamics to Global Consequences." *Journal of Ecology* 97 (5): 840–50. https://doi.org/10.1111/j.1365-2745.2009.01529.x.

Sun, Zhe, Jianjun Zhao, Hongyan Zhang, et al. 2025. "Predicting the Start of the Growing Season in Boreal Forest Under High and Low Emission Scenarios." *Earth's Future* 13 (8): e2024EF005622. https://doi.org/10.1029/2024EF005622.

Tang, Yongming, Steve Rumbold, Rich Ellis, et al. 2019. "MOHC UKESM1.0-LL Model Output Prepared for CMIP6 CMIP Historical." Version 20230220. With Colin Jones, Matthew Mizielinski, Mark Elkington, et al. Earth System Grid Federation. Application/x-netcdf. https://doi.org/10.22033/ESGF/CMIP6.6113.

Van Vuuren, Detlef P., Elke Stehfest, David E. H. J. Gernaat, et al. 2017. "Energy, Land-Use and Greenhouse Gas Emissions Trajectories under a Green Growth Paradigm." *Global Environmental Change* 42 (January): 237–50. https://doi.org/10.1016/j.gloenvcha.2016.05.008.

Zelinka, Mark D., Timothy A. Myers, Daniel T. McCoy, et al. 2020. "Causes of Higher Climate Sensitivity in CMIP6 Models." *Geophysical Research Letters* 47 (1): e2019GL085782. https://doi.org/10.1029/2019GL085782.

Zhang, Xiaoyang, Mark A. Friedl, Crystal B. Schaaf, et al. 2003. "Monitoring Vegetation Phenology Using MODIS." *Remote Sensing of Environment* 84 (3): 471–75. https://doi.org/10.1016/S0034-4257(02)00135-9.

https://doi.org/10.5194/egusphere-2026-1710
Preprint. Discussion started: 8 April 2026

Zhang, Xiaoyang, Mark A. Friedl, Crystal B. Schaaf, and Alan H. Strahler. 2004. "Climate Controls on Vegetation Phenological Patterns in Northern Mid- and High Latitudes Inferred from MODIS Data." *Global Change Biology* 10 (7): 1133–45. https://doi.org/10.1111/j.1529-8817.2003.00784.x.

Zheng, Lei, Youcun Qi, Yijie Wang, Jie Peng, and Zhangcai Qin. 2022. "Calibration and Validation of Phenological Models for Biome-BGCMuSo in the Grasslands of Tibetan Plateau Using Remote Sensing Data." *Agricultural and Forest Meteorology* 322 (July): 109001. https://doi.org/10.1016/j.agrformet.2022.109001.

https://doi.org/10.5194/egusphere-2026-1710
Preprint. Discussion started: 8 April 2026

Appendix

A.1 GS-P description

GS-P is a probabilistic, temperature-based model designed to estimate the SGS and EGS through a two-step procedure that combines machine-learning classification with inference based on probabilistic thresholds.

In the first step, GS-P estimates the daily likelihood that a given grid cell is in a GS or non-GS state based the following temperature-based features: the 14-day moving average of the daily temperatures (T_{14d}), the associated regression slope (∇T_{14d}) and moving standard deviation (σ_{14d}), and the 60-day moving average (T_{60d}). This is achieved through a binary classification model, calibrated using the XGBoost algorithm (Chen and Guestrin, 2016; see Table A1 for hyperparameters). XGBoost constructs an ensemble of decision trees iteratively, where each successive tree is trained to reduce the residual error of the previous ensemble. For each day and grid cell, the model outputs a raw score that represents the model's confidence in the GS state. This raw score is then transformed through a logistic function, producing a daily probability that indicates the likelihood of being in a GS state. The probabilistic output accounts for non-linear interactions among the input features ($T_{14d}, \nabla T_{14d}, \sigma_{14d}, T_{60d}$) allowing the model to capture both absolute temperature thresholds and dynamic patterns in temperature changes.

In the second step, SGS and EGS are inferred from the daily GS probabilities by introducing two probability thresholds (p_{SGS} and p_{EGS}), which define the probability levels at which transitions between non-GS and GS are considered to occur. For each year, the model identifies the temporal window corresponding to that year and considers the probability curve for each node. The convexity of the probability curve at each node is assessed using the smoothed daily probabilities (5-day moving average). A second-degree polynomial is fitted to the smoothed probability curve over the year, and the coefficient of the quadratic term is used to classify the curve shape: a positive coefficient indicates a convex curve, while a negative coefficient indicates a concave curve. Convex and concave shapes are typical (but not exclusive) of the NH and SH seasonal cycles, respectively. Using this information, candidate SGS and EGS days are constrained relative to the peak probability: in the NH, SGS occurs before the peak and EGS after it, while in the SH, the ordering is reversed. The final SGS and EGS for

https://doi.org/10.5194/egusphere-2026-1710
Preprint. Discussion started: 8 April 2026

each node and year are selected as the earliest and latest valid candidate days, ensuring that each year has a single, consistent start and end of the growing season, even in cases where multiple threshold crossings occur due to variability or noise in the probability signal.

Table A1. Hyperparameters for the XG-Boost algorithm. The R package xgboost was used (https://cran.r-project.org/package=xgboost).

General parameters	
objective	binary:logistic
eval_metric	logloss
max_depth	5
tree_method	hist
nrounds	100

A.2 Parameter calibration details

The loss function for GS-Lin2 and GS-BC2 is based on the RMSE, which is defined as:

$$RMSE(Y,\hat{Y}) = \sqrt{\frac{\sum_{i\in I}(Y_i-\hat{Y}_i)^2}{n}} \quad \text{(A7)}$$

Where $Y = (Y_i)_{i\in I}$ are the MODIS observation, $\hat{Y} = (\hat{Y}_i)_{i\in I}$ are the values predicted by the model, and $I = \{1,\dots,n\}$ is the number of values for which the model actually predicts dates for SGS and EGS. The RMSE can be found for both SGS and EGS, comparing the observed and the predicted dates. For every model, the loss function is defined as the median of the two RMSEs, with a penalisation to account for the number of times $\tilde{n}$ in which the model failed in producing a viable result:

$$\lambda(Y,\hat{Y}) = \left(1+\frac{\tilde{n}}{N}\right)^2 \cdot \frac{RMSE_{SGS}(Y,\hat{Y})+RMSE_{EGS}(Y,\hat{Y})}{2} \quad \text{(A8)}$$

Where $N = n + \tilde{n}$. In other words, when the GS model always succeeds in predicting SGS and EGS over a certain dataset, $\tilde{n} = 0$ and $n = N$ and no penalization is applied. On the other hand, the more often the model fails, the higher $\tilde{n}/N$ will be, which makes the corresponding $\lambda(Y,\hat{Y})$ to increase (penalisation). In any case, to avoid parameters being calibrated over too few points, $\lambda(Y,\hat{Y})$ is set to 1000 d when $\tilde{n}/N > 0.2$.

The joint RMSE for SGS and EGS is required by the GS-Lin2 model, as this model requires the SGS to start either before or after the EGS, depending on the hemisphere (and therefore, at a global scale, it is not possible to calculate the SGS without calculating the EGS, and vice versa). On the other hand, GS-BC2 calculate the loss function that minimizes the RMSEs separately, i.e. the loss function does not consider the median RMSE of SGS and EGS, but just one of the two, as the parameters $t_{1,SGS}$ and $t_{1,EGS}$ naturally lead to an event separation. Instead, GS-Lat depends only on geographical coordinates,

https://doi.org/10.5194/egusphere-2026-1710
Preprint. Discussion started: 8 April 2026

and not on temperature. Therefore, this model always predicts SGS and EGS by design, and in this case the penalization would be null, which makes ordinary least squares equivalent to finding the minimum of the defined loss function.

In GS-P, probability thresholds are inferred separately for SGS and EGS on a grid node level. For each grid cell, observed SGS and EGS dates derived from MODIS are first matched to the corresponding daily GS probabilities produced by the classifier. These probabilities form node-level samples of transition probabilities $\{P_{i,t}\}_{t=1}^{n_i}$, with i indicating the grid-node, and n_i the number of observations in the node i. Threshold inference is performed within each KG climate class using a hierarchical Bayesian Normal-Normal model with empirical shrinkage (Gelman et al. 2025). This model assumes that node-specific thresholds p_i ($p_{i,SGS}$ or $p_{i,EGS}$) are drawn from a KG class-level distribution $N(\mu, \tau^2)$, where μ is the KG class mean threshold and τ^2 is the between-node variance. Observed node-level probabilities are modelled as $N(\mu_i, \sigma^2)$, where μ_i is the within-node mean, and σ^2 is the pooled within-node variance. The node-specific thresholds $p_{i,SGS}$ and $p_{i,EGS}$ are obtained via shrinkage as precision-weighted average of the sampled mean of the probabilities ($\bar{P}_i$, estimator of μ_i), and the KG class-level mean:

$$p_i = w_i \bar{P}_i + (1 - w_i)\mu \quad \text{(A9)}$$

Where the weights w_i are defined by:

$$w_i = \frac{\tau^2}{\tau^2 + \sigma^2/n_i} \quad \text{(A10)}$$

With σ^2 estimated from pooled within-node variance, and τ^2 from the method of moments.

In practice, when within-node variance is high (or there are few observations at node-level), $w_i \to 0$ and consequently $p_i \to \mu$. On the other hand, when within-node variance is small, $w_i \to 1$ and $p_i \to \bar{P}_i$. Here, $p_{i,SGS}$ and $p_{i,EGS}$ indicated node-level thresholds; for readability, the subscript i is omitted elsewhere in the text, and these thresholds are indicated as p_{SGS} and p_{EGS}.