

Reviewer #1

General comment

This manuscript presents a relevant and timely contribution to global forest phenology modelling. The authors introduce GS-P, a probabilistic temperature-based model designed to estimate the start and end of the growing season of global forest ecosystems from 1850 to 2100. The study combines MODIS phenology data, ERA5 reanalysis, and UKESM1-0-LL climate simulations, and compares the proposed model with alternative temperature-based approaches and a latitude-based benchmark.

The manuscript is scientifically interesting, generally well structured, and clearly relevant to the readership of Biogeosciences.

The main strengths of the paper are the global perspective, the long temporal coverage, the comparison with benchmark models, and the explicit treatment of prediction uncertainty. The GS-P model appears particularly useful as a flexible and interpretable framework for identifying temperature-driven phenological transitions and regions where temperature alone may provide limited explanatory power. My recommendation is for a minor revision. The manuscript is suitable for publication after some revisions aimed at improving clarity and moderating some ecological interpretations.

We thank the reviewer for the supportive comments and for the useful suggestions. We revised the paper following the provided comments as much as possible, while accounting for the major changes required by reviewer #2.

Responses are in red. Cited parts of the revised article are in green.

Detailed comments

Abstract

L 10–25: The Abstract is clear and effective, but some statements could be slightly softened. In particular, the references to “ecosystem vulnerability”, “carbon cycle modelling”, and “climate–ecosystem disequilibrium” may suggest stronger ecological conclusions than are directly demonstrated by a simple temperature-based phenological model.

I suggest rephrasing these claims as potential applications or interpretations. For example, rather than stating that the model identifies ecosystem vulnerability, the authors could state that GS-P could identify regions where phenological predictions are more uncertain and where additional ecological investigation may be needed to preserve ecosystem stability.

Following also the comments of reviewer #2, the paper was restructured by offering a more nuanced interpretation of GS-P features. Particularly, in the abstract, we removed references to the interpretation, limiting the discussion of model parameters to the discussion section (section 4.3).

Introduction

L 38–55: The discussion of the different definitions of growing season is useful and important. The distinction between phenological, productive, meteorological, and biological growing seasons should be carried more consistently into the rest of the manuscript. The authors correctly state that meteorological growing season represents potential vegetation activity rather than actual growth. At the end of the Introduction, add one sentence clearly defining what GS-P estimates, for example:

“Hereafter, GS-P estimates are interpreted as temperature-conditioned estimates of phenological growing-season transitions, calibrated against MODIS phenological markers.”

We thank the reviewer for the suggestion. The sentence was modified and added to the last paragraph of the introduction:

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. **As such, the output from these models should be effectively interpreted as temperature-conditioned estimates of GS transitions (*meteorological season*), calibrated against MODIS-derived phenological markers (*phenological season*).** The ultimate outcome of this study...

L 78–88. The motivation is convincing, but the sentence suggesting that the product may reduce uncertainties in terrestrial carbon-cycle models and support forest management should be slightly moderated. Use wording such as “may contribute to” or “may provide useful constraints for”, rather than implying a direct demonstrated reduction of uncertainty.

The sentence was moderated as suggested:

Particularly, because our framework is primarily data-driven, the resulting product can provide a long-term, relatively independent benchmark for terrestrial ecosystem and vegetation models.

Data and methods

L 105–113: The use of UKESM1-0-LL is clearly described. However, since future projections presented in the paper rely on one ESM, this limitation should be stated explicitly in the Methods and recalled in the Discussion or Conclusions. Add a sentence noting that the future projections represent GS-P responses under UKESM1-0-LL forcing and do not sample the full CMIP6 inter-model range.

We thank the reviewer for the remark. The following was added to the discussion (section 4.1):

The relatively large shifts projected here may be also caused by the marked warming projected by UKESM1-0-LL, which is characterized by one of the strongest warming trends in the CMIP6 constellation (Zelinka et al. 2020). While the choice of this model was made specifically to test GS-P responses to pronounced climate warming,

a limitation of this study is that it does not account for the full inter-model range of climate projections – as a matter of fact, GS simulations driven by other ESM forcings may result in smaller, or otherwise different, SGS and EGS shifts.

Further warnings will be added to the open-source repository once the final version of the dataset is uploaded.

L 185–189: The calibration/validation split is reasonable, but it is based partly on random sampling. Because phenological and climatic data are spatially autocorrelated, the authors should briefly acknowledge that this validation mainly tests predictive skill within the sampled spatial–climatic domain. I do not consider additional validation essential for publication, but the manuscript would benefit from a short comment on the possible value of future spatial block validation or forest-type-specific validation.

Following a comment of the reviewer #2 we modified the dataset split, and repeated calibration and validation. The manuscript is updated accordingly to the new results. The dataset split is explained in section 2.3

Of all available years (2001–2023), the last five years (2019–2023) are withheld for validation, while the rest (2001–2018) are used for calibration. Moreover, to further validate the models spatially, the globe is divided into 12 uniform blocks (four longitudinal and three latitudinal splits), among which four are randomly withheld for validation every six years.

L 240–257: The uncertainty approach is a strength of the paper, but the interpretation of the prediction intervals should be clarified. The current description explains how the uncertainty estimates are derived, but not whether they are empirically calibrated against validation observations. Add a short sentence clarifying whether the reported intervals should be interpreted as empirical error envelopes rather than fully probabilistic confidence intervals. If available, report the empirical coverage of the intervals on the validation dataset.

We verified that the empirical coverage of the prediction intervals against the validation observation, and found it to be approximately 95%. By definition, because the uncertainty intervals are derived from the empirical distribution of bootstrap predictions, they represent empirically derived prediction envelopes rather than formal parametric confidence intervals. As such, we prefer to retain the current formulation of the uncertainty approach to avoid unnecessarily expanding the methodological description.

Results

L 321–331. The model performance is clearly presented. GS-P achieves strong global improvements in MAE, but the manuscript also shows that performance is substantially better in continental than in temperate climates. The sentence stating that MAE values of 3.0–7.6

days can be considered “negligible for most applications” should be softened. Such errors may be small for global or hemispheric trend analyses, but they may still matter for local phenological studies, frost-risk assessment, or species-specific applications.

Suggested revision: Replace “negligible for most applications” with “negligible for many large-scale applications”.

Done.

L 336–352: The discussion of spatial bias patterns is useful. I suggest adding one sentence emphasizing that regions with large bias or large error spread should be treated cautiously by users of the dataset, especially where other factors such as water availability, monsoon dynamics, evergreen phenology, or topography may influence phenological timing.

We thank the reviewer for the remark. However, it is implicit that areas with larger mean bias or error spread are to be treated carefully. As such, we preferred to avoid adding this sentence.

Discussion

L 480–491: This section appropriately notes that temperature alone may not provide sufficient explanatory power in the Southern Hemisphere and that additional drivers may be important. This is an important point and should be made slightly more visible, perhaps also in the Conclusions.

Add a concise “recommended use” statement: GS-P is expected to be most reliable where temperature is the dominant seasonal control, while outputs should be interpreted with more caution in regions with weak temperature–phenology coupling.

Following also remarks of the Reviewer #2, we expanded the section related to additional drivers and to the Southern Hemisphere (please see reply to RC#2). Regarding the “recommended use” statement, we thank the reviewer for the suggestion. Such statements will be added to the description of the repository once the final version of the dataset is uploaded.

Lines 503–521. The ecological interpretation of GS-P parameters is interesting and potentially valuable. However, the wording should be slightly moderated. The model parameters are inferred from temperature–phenology relationships calibrated against MODIS observations; therefore, they should not be presented as directly capturing physiological features of forest ecosystems. Rather, they reveal statistically derived patterns that are consistent with known ecological and phenological expectations, such as stronger temperature control of spring phenology than autumn phenology and clearer temperature–phenology coupling in continental climates. I suggest replacing expressions such as

“physiological features captured by the model” with “ecologically interpretable phenological patterns inferred by the model” or “patterns consistent with known physiological and ecological controls.” “Although GS-P cannot directly resolve physiological processes, its parameters provide ecologically interpretable information on temperature–phenology relationships across climate classes.”

We thank the reviewer for the suggestion. The section of the discussion was moderated, and the specification regarding the interpretation was added to the discussion section (section 4.3). Please see also the reply to RC#2 for more details.

Lines 522–530. The GS-flipping discussion is one of the most interesting parts of the manuscript, but it also requires careful framing. The authors already note that it may reflect either climate–ecosystem disequilibrium or a limitation of the current framework.

Suggested revision:

Present GS-flipping primarily as a model diagnostic indicating extrapolation beyond well-established temperature–phenology relationships. The ecological interpretation can remain, but should be framed as a hypothesis requiring further testing.

We thank the reviewer for the comment and agree with their interpretation. The GS-flipping feature was relatively relevant in the old version of the manuscript. The current configuration of GS-P considers the photoperiod among the model predictors. This improved the model’s performances and, consequently, reduced GS-flipping instances to only very few nodes. As such, the reference to the GS-flipping feature was removed from the manuscript.

Conclusions

L 533–543. The Conclusions are generally clear, but the projected extension of the growing season should be explicitly linked to UKESM1-0-LL forcing. Change wording such as “under climate warming” to “under UKESM1-0-LL climate forcing” or “in the UKESM1-0-LL-based projections”.

We thank the reviewer for the suggestion. Explicit mention to UKESM1 has been added to the conclusions:

In this study, we developed GS-P, a probabilistic, temperature-based model designed to estimate SGS and EGS of generic forest ecosystems at a global scale. Estimates of GS transitions were produced from 1850 to 2100, using bias-corrected, downscaled UKESM1-0-LL climate forcing.

L 537–543. The statements that GS-P identifies biological features, climate–ecosystem mismatch, and priority zones for conservation are interesting but should be slightly moderated. Use “suggests”, “may indicate”, or “highlights regions that need further ecological investigation”, rather than more definitive language.

We removed the statements related to GS-P interpretability from conclusions – as such, these modifications cannot be made.

Line 557. The data availability statement is useful, but it only mentions the GS-P output.
Suggested revision: Clarify whether the code used for preprocessing, calibration, prediction, and figure generation will also be made available.

We specified that the code will be made available in Git.

Language and grammar corrections

We thank the reviewer for the corrections. They were integrated into the manuscript.

The manuscript is generally well written, but several small language corrections are needed.

L 42: “which aspect are covered” → “which aspects are covered” or “which aspect is considered”.

L 109: “primarily” → “primarily”.

L 130: “generic forests ecosystems” → “generic forest ecosystems”.

L 149: “comparatively to others” → “compared with others” or “relative to others”.

L 242: “To compare the performance the different GS model” → “To compare the performance of the different GS models”.

L 264: “comparatively to D” → “compared with D” or “relative to D”.

L 266: “only warm spells of only 3 days” → remove one “only”.

L 484: “substantial higher number” → “substantially higher number”.

L 514: “species compositions” → “species composition”.

L 550: “synchronous climate-GS regimes transitions” → “synchronous transitions between climate and GS regimes”.