Dear Shilong Ren,

Thank you for your community comment 1 and your interesting question to the initial submission of our manuscript. We are pleased to provide a strongly revised version, based on your comment and on the comments of the reviewers 1 and 2. All changes are highlighted in yellow, whereas light yellow indicates shifted but unchanged text. Please find our detailed responses to your comment here below (**bold** text).

Best regards,
Michael Meier, Christof Bigler, and Isabelle Chuine

* * *

Community comment 1
(https://doi.org/10.5194/egusphere-2025-460-CC1)

The authors developed a new and systematic model of the senescence process of woody plants and analyzed the impact of data quality on the simulation of autumn phenology. The research perspective is unique. After previous revisions, I think there are no issues with the article's structure and writing quality. I have only one question: According to the multi-model comparison results in the article, it is not difficult to find that the DP3 model does not outperform the PAI and DM2 models in terms of simulation accuracy. It seems that the improvement significance of the new model is not significant. Could it be that the model process is too complex and has too many parameters?

We have now included the results for the DP3 model calibrated with one and with two stages of leaf senescence (i.e., with the stage when 50% of the leaves have turned color or fallen, LS_{50}, as well as with LS_{50} and the stage when 100% of the leaves have turned color or fallen, LS_{100}). These two calibrations led to different results. For example, the young leaf phase lasts 41 days when calibrated with LS_{50} and LS_{100} but only 1 day when calibrated with LS_{50} only. This illustrates a compensating effect between the different model parameters (i.e., different parameter sets yield similar results), which we now discuss in L441–446). However, while the probability for such compensating effects arguably increases with the number of parameters, it should be irrelevant for the accuracy of the predictions, provided that these remain in the space of the calibration conditions. This accuracy may only suffer by a high number of parameters when the calibration algorithm cannot handle them. Here, we selected the algorithm generalized simulated annealing, which has been used to successfully calibrate models with up to 30 parameters (Xiang et al. 2013, https://doi.org/10.32614/RJ-2013-002). Moreover, we carefully tuned this algorithm to the models (Sect. S2.2), why we are confident that the complexity of the model did not have any adverse effects on the accuracy of the predictions.