Dear Reviewer 2,

Thank you for your positive response to the initial submission of our manuscript. We are pleased to provide a strongly revised version, based on your comments and on those of reviewer 1 and of Shilong Ren (community comment 1), including our own further revisions. All changes are highlighted in yellow, whereas light yellow indicates shifted but unchanged text. Further, we provide our detailed responses to your comments here below (**bold** text). Note that references here below without DOI also appear in the manuscript, where they are listed including DOI.

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

* * *

Reviewer 2
(https://doi.org/10.5194/egusphere-2025-460-RC2)

General Comments

This study presents an interesting and innovative approach to modeling leaf senescence, which remains a challenging process to simulate. The work offers two key contributions:

Unlike previous process-based models that focus solely on leaf senescence, the DP3 model attempts to represent the entire leaf development process from spring to autumn (that means from leaf unfolding to leaf senescence).

The authors analyze the influence of leaf senescence data quality on model performance, which often overlooked in modeling studies.

While the proposed model introduces a novel structure, the results indicate that it does not yet simulate leaf senescence dynamics well. The authors attribute this primarily to data quality limitations. However, given the complexity of leaf development processes (from unfolding to senescence), model performance may also depend on how well these processes are represented. Phenology data derived from camera observations (e.g., PhenoCam) are less susceptible to observer bias and sampling uncertainty compared to traditional ground observations. Have the authors considered using such datasets to further evaluate the model structure?

While we thoroughly discuss the representation of the processes in the model formulation (L404–418, L485–494, L521–530) as well as the uncertainty in visually recorded phenology data (L511–519, L531–541), we did not consider to reevaluate the model with new data from PhenoCams. Although this is a very compelling idea and would likely yield valuable results, we have no such data right now to try this and would be very interested in accompanying such a new study with researchers who would have extended enough time series of leaf senescence dates evaluated with images.

The hypothesis that aging and stress drive leaf development is compelling. A discussion comparing this approach with more conventional growing-degree-day-based models would strengthen the manuscript. For instance, what are the advantages of using aging and stress as drivers instead of accumulative growing season temperature?

We wrote a new paragraph to discuss this (L404–418) and tried to include this thought in the introduction to newly derived hypotheses from the DP3 model (L427) to discuss it later (L454–464).

Interestingly, the results highlight cold, daylength, and dry stress as key drivers—similar to the Growing Season Index (GSI) model, which relies on minimum temperature, photoperiod, and vapor pressure deficit (VPD). Did the authors test VPD as an alternative drought stress indicator?

Unfortunately, we did not test any alternative drought indices. However, we now discuss this shortcoming in section 4.4 (L545–550).

Specific Comments

Line 43, 105: Please clarify the definition of leaf senescence. Autumn phenology typically distinguishes between leaf coloring and leaf fall as separate stages. How is senescence defined in this study based on both events?

While we use leaf senescence as collective term for leaf coloring and leaf fall (now stated accordingly in L44–45), we based our study on the autumn phenology stages BBCH95 and BBCH97 for pome and stone fruit according to Meier (2018, https://doi.org/10.5073/20180906-074619) as now specified in L121–122.

Figure 1C: The delayed leaf senescence at higher latitudes appears counterintuitive, as senescence usually occurs earlier in such regions. Could the authors provide insight into possible causes for this pattern?

The regression through the function geom_smooth in the R package ggplot2 was calculated separately to each response variable (i.e., average LS_{50} and average LS_{100}) as well as separately to each explanatory variable (latitude, longitude, and elevation. Thus, a positive relationship emerged between both LS_{50} and LS_{100} and latitude probably because the more northern sites are generally lower elevated. This is misleading, as you have pointed out, and we have corrected it accordingly. In the revised version of the manuscript, we fitted a linear regression to combined latitude, longitude, and elevation for each response variable (Sect. S1.1.2). These regressions indicate a negative relationship between the response variables and each explanatory variable. Figure 2c (former Figure 1c) was adjusted accordingly by plotting the results of these regressions for each explanatory variable, while keeping the other explanatory variables constant (i.e., set to the mean).

Table 1:

Please include details on the spring phenology (LU, leaf unfolding) data.

The additional information was included.

The dataset combines observations from PEP725 and SPN. Were these collected using the same protocols? If not, how might protocols' differences affect model performance? Have the authors tested the model using only one dataset to assess potential improvements?

We had no access to the precise protocols used to collect the data. As these protocols were established by different institutions from different countries, they likely differ (Menzel, 2013, https://doi.org/10.1007/978-94-007-6925-0_4). However, the same stages were visually observed among countries (i.e., the stages BBCH15, BBCH95, and BBCH97 according to Meier, 2018, https://doi.org/10.5073/20180906-074619). While we did not use data from only one country, the example script we provided together with the code for the DP3 model (Meier, 2025, https://doi.org/10.5281/zenodo.14749339) runs on data from only three sites. In consequence, the accuracy of the predictions for both the observations in the calibration sample and validation sample is considerably improved. This emphasizes our suggestion that the heavy noise in the used data blur the signal of leaf senescence. Thus, comparing the DP3 model with current models based on observations that do not contain any sudden changes in the mean (Auchmann et al., 2018) is an important way forward, which we now suggest in lines 576–580.

Line 121: Please briefly describe the E-OBS dataset.

We now do so (L139–141).

Lines 124–125: Could the authors elaborate on the temperature correction method applied?

We had done so in the lines 127–130 of the original manuscript, but probably did not emphasize this enough. We now restructured the paragraph a bit, such that the temperature correction method is now easily identified (L143–148).

Line 135: Which remote sensing dataset was used?

Here, we referred to the remote sensed CO_2 dataset. However, this was unclear, so we recited the dataset (L156).

Line 139: Since LAI and CO_2 concentration are provided as monthly data, how was daily photosynthetic activity derived?

These data were combined with daily values of surface shortwave down welling radiation, day length, and mean temperature. We now have clarified this in L160–161.

Line 141: The Keetch-Byram Drought Index (KBDI) was selected as the drought metric. Were other indices (e.g., SPEI, PDSI) tested? If so, how did they compare?

Unfortunately, we did not test any other drought indices. Considering the many drought indices there are, such a comparison would have inflated our manuscript too much. However, we totally agree with you that such a comparison would certainly be very valuable and believe that it would yield an entire study by itself. Nevertheless, we now briefly discuss this in section 4.4 (L545–550).

Line 152: Does "several" refer to 34 sites? If so, please specify for clarity.

Yes, in the end, we constructed and tested 34 formulations. However, this number of formulations is a result rather than a component of the method applied. Therefore, we mention it in the first line of the results section (L328). However, in order to avoid confusion, we simply omitted the word «several» in the method section (L173).

Lines 200–204: The parameters *a* and b0 are set to 0.01 h. Could the authors justify this choice?

These are examples. The calibrated values are listed in Table 3. We now made this clearer in the text (L224).

Lines 218–219: Please define "extreme conditions" (e.g., hottest temperatures >30°C, coldest below a certain threshold).

Rather than using a threshold to identify these conditions, we selected the site-years that contained the hottest 10-day period during the growing season observed in the dataset. We did so, because we wanted to select exactly 250 site-years. We have now specified this in lines 240–241.

Figure 4 (3rd iteration): Does *f* represent h(x)? If so, please clarify in the caption.

Yes, it does. We have now corrected this and revised the figure (now Fig. 6) completely.

Table 2:

dm→o: Should this be interpreted as the simulated transition timing from mature to old leaves?

Yes, it should. We have now corrected the definition (Table 2).

SAging,I / SStress,i: Do these states accumulate since LU (leaf unfolding)?

No, these states are the accumulated corresponding rates since the transition from young to mature leaf. We have now clarified this in the definition column of Table 2.

Line 380: The authors associate cold stress with spring frost events. Could the importance of cold stress after midsummer also be examined?

Not directly. An additional assessment would be necessary to examine cold stress accumulated before and after summer solstice. Rather than including such an additional assessment, we now included some results regarding the relative importance of cold stress during senescence (i.e., the period from senescence induction to leaf senescence; L372–376; Fig. S3; Tables S5–S8) and further discussed effects of cold stress (L454–478).

Line 388: "This maybe" → Please revise for clarity (e.g., "This may be due to...").

We modified the sentence as «This may be explained by unrealistic model formulations, poor model calibrations, and noisy data to drive and calibrate the models, all of which we discuss here below» (L480–484).

Line 400: The sentence structure could be improved for readability.

We adjusted the sentence structure (L497–500).