Revision of Manuscript egusphere-2025-1534

In response to Reviewer 1 and the editorial request, we have revised the manuscript and Supplementary Material as follows:

1. *Additional falsification analysis and new figures in Supplementary Material*
   - Performed an additional falsification analysis in which AWC-scenario selection is based only on the 2000–2017 period.
   - Added the corresponding supplementary figures (for beech and spruce) that mirror the structure of Figs. 3 and 4 but restrict the comparison of observed vs. simulated mortality to 2000–2017.
   - Described these results in the revised Results section and expanded the discussion of scenario-selection uncertainty and period dependence.
2. *Clarified interpretation of pre-2018 mortality and model behavior*
   - Revised the Results and Discussion to clarify that pre-2018 mortality peaks are not purely stochastic but are linked to threshold-like model responses to moderate-to-severe drought conditions, while also highlighting cases where the model overestimates mortality.
   - Clarified the role of the drought index and medium drought conditions in triggering mortality under low AWC settings.
3. *Textual and structural clarifications in the main manuscript*
   - Corrected typographical and consistency issues pointed out by the reviewer 1
   - Regularized the Methods section numbering (including correcting the missing Section 2.2 and renaming Section 2.2.4 to "Simulation rationale and settings") and moved the statement about the model-intrinsic nature of the results to a more prominent position at the beginning of that section.
   - Added a clarification that the residuals were approximately Gaussian, justifying the Gaussian error assumption.
   - Implemented wording refinements suggested by the reviewer 1
4. *Clarification of PIC fractions and stochastic/background mortality*
   - Revised the caption and explanation of the relevant supplementary figure showing fractions of mortality attributed to PIC predisposing and inciting stress factors
5. *Revisions to the Supplementary Material and cross-references*
   - Renumbered all supplementary figures and tables consecutively (Fig. S1, S2, …; Table S1, S2, …), independently of section numbering.

- Updated all references to supplementary figures and tables in the main text and Supplement to match the new numbering.

6. *Data and code availability*

    - Added and highlighted the DOI of the dataset (Marano and Bugmann, 2025) at its first mention and in the "Code and data availability" section.

A detailed, point-by-point response to all comments from Reviewer 1 is provided in the following section of this document.

Author responses to Reviewer 1 of egusphere-2025-1534

We thank the reviewer for the in-depth assessment of our work. We provide a point-by-point response below in red.

I highly appreciate the efforts the authors have taken in response to my first review. All in all, the additional analyses and textual refinements have improved the robustness of the manuscript which I deem almost ready for acceptance.

Yet, one relatively important aspect needs further consideration which refers to the selection of AWC-scenarios based on r^2 and MAE. In particular, the authors now provide more detailed insights into the behavior of r^2 and MAE in the period prior to 2018, indicating – as suspected – a lower fit between simulated and observed mortality rates. However, they explain this by stochasticity of mortality prior to 2018 which I do not see supported by some of the display items. That is, some of the mortality peaks prior to 2018 in fact seem to be caused by drought, wherefore a satisfying model fit in that period is desirable, too.

While this aspect does not question the overall mortality implementation at all (which I personally think is a strong point of the manuscript since not being based on calibration but model-internal mechanisms) it does have an impact on the presented results. To make this clear early on: I am not asking for major revisions of the results section. But I would like to see a few more aspects of falsification and visualization of uncertainty in the supplementary alongside a brief discussion/evaluation of these items in the results and discussion. This basically includes two figures comparable to Figs. 3 and 4 but only for the calibration period 2000-2017. This would clearly visualize, how a different period used for the identification of the best AWC-scenarios would affect the selection and consequently the simulated mortality rates. The authors can then briefly mention these effects in the corresponding results sections (lines 364-370 for beech and 438-444 for spruce) and elaborate the discussion in lines 479-486 and/or 572-576. Doing so, allows readers to fully capture the effects of period selection on model selection and the related results.

To make clear, why I think this point is so important I would like to bring up a hypothetical scenario: Imagine, the study were done in 2017 and you would only have the observational data available until that year. Back then, the selected AWC-scenarios would be different (as indicated by Fig. S 3.3.1). It would be interesting to see how these other models predict the mortality after 2017, i.e. outside the period that was used as a baseline for selecting these models as best fit. This hypothetical scenario stands example for the current situation, where we do not know what will happen in the future. So examining this hypothetical example would help to assess the uncertainty related to the AWC-scenario selection approach and related interpretations.

We thank the reviewer for these insightful and constructive reflections and comments. We agree that this evaluation strengthens the transparency and credibility of the scenario selection process, particularly under the hypothetical condition where the future (post-2017) is unknown. Therefore, we have conducted a falsification analysis as requested, using only the 2000–2017 period for model evaluation and scenario selection. The results of this analysis are now included in the Supplementary Materials (Figures S10 and S3). These figures mirror the format of Figures 3 and 4 but limit the observed vs. simulated mortality comparison to the pre-2018 period. This approach, together with Figures S 3.2.1.1 and S 3.2.2.2 respectively, allows for a clearer visualization of how the AWC scenario selection would differ if only earlier data had been available.

For beech (Fig. S10), this led to the selection of scenarios 1 and 4 (Fig. 10, D-E), both representing very low AWC conditions (AWCmean = 7.5 cm; AWCmin = 7.5 and 15 cm, respectively; see red boxes in Figure S10, A). This is consistent with the model's stronger drought sensitivity under low AWC (Figure S8), with mortality peaks occurring in known drought years (e.g., 2003–2004 and 2012), as shown in Figure S9, A. We acknowledge the reviewer's point that these peaks reflect a response to drought rather than stochasticity, and we have revised the manuscript (lines 371-376) accordingly to clarify that the observed mortality prior to 2018 is not random but linked to threshold-like model behavior in response to moderate-to-severe drought conditions. However, there are also years such as 2015–2016 (scenario 4) where the model simulates mortality spikes not supported by observations (Figure 1, main text). These may result from an overreaction to the simulated drought intensity and combined with background mortality events, in the absence of observed mortality peaks (lines 376-378).

For spruce (Fig. S13), the scenarios selected under the 2000–2017 window were scenarios 29 (AWCmean = 15 cm, AWCmin = 12.5 cm) and 41 (AWCmean = 17.5 cm, AWCmin = 15 cm) (see red boxes in Figure S13, A and panels D-E). These closely align with the full-period selection (where scenario 42 was preferred over 41), suggesting a more stable scenario ranking for spruce irrespective of the calibration period. The performance metrics for these scenarios (Adj R^2 = 0.258 and 0.212; MAE = 0.733% and 0.631%, respectively) indicate reasonable consistency despite slightly lower observed mortality and weaker drought signals in the earlier years. Similar to beech, stronger drought years tend to trigger mortality more readily under low AWC, although in some cases—such as 2003–2004, and especially 2010–2012 and 2015–2016—the model overshoots, simulating mortality peaks not evident in the observations. This behavior reflects persistent medium drought index values (DI > 0.1–0.15) during those years (Figure S12, A), which drive the model response, particularly under low AWC settings.

Overall, this additional analysis confirms the qualitative consistency of model behavior across periods: low AWC scenarios lead to early and often elevated drought-induced mortality for both species. However, it also highlights the sensitivity of scenario selection to the calibration window, particularly for beech. We now explicitly discuss this in the revised Results section (lines 371-378 for beech; lines 451–455 for spruce) and in the Discussion section (lines 493–497 and 592–595). Notably, spruce scenario rankings were more temporally stable than those of beech, pointing to species-specific differences in how robustly model selection captures mortality dynamics across time.

We agree with the reviewer that model evaluation under both moderate and extreme drought conditions is essential, especially if the model is to be used for future projections. Our additional analysis therefore contributes to a more balanced assessment of scenario uncertainty, which does not only focus on the post-2018 period.

Apart from this, I only have a few minor points as outlined below.

Line 16: 'four' hypotheses is probably a legacy from the initial submission. I guess it should be three

We have corrected this at line 16.

Line 37: it seems the word 'of' is missing in this sentence: understanding, forecasting and managing 'of' forest resistance

We have corrected this at line 37.

Lines 111 and 122: it is justified to only select the sites from Knapp et al. (2024) but I nevertheless suggest to reflect in the text why northeastern Germany is basically empty, since readers may wonder (as I did) why that is the case.

We have added a consideration about the lack of sites in N-E Germany at lines 113-116.

It seems that section 2.2 is missing, please adjust (I guess section 2.3 should be 2.2)

We apologize for this oversight. We have corrected all numbering accordingly.

Lines 232 and 234: BGB has not been introduced before. Or do the authors refer to BGR which was mentioned in line 225? Please clarify.

We apologize for this glitch; we referred to the BGR. We have corrected this at lines 235 and 237.

Lines 256-260: I wonder whether the authors want to find a more prominent spot for mentioning this important point, e.g. at the beginning of section 2.3 (which should be 2.2. I guess) to stress that the results represent model-intrinsics and not empirical fits which I personally evaluate as a major strength of the presented approach.

We have moved this consideration to the beginning of section 2.2.4, which we renamed to "Simulation rationale and settings" (lines 214-219).

Line 282: I understand from your reply that your data were quasi-normal distributed wherefore you assumed a Gaussian fit. Just for the sake of clarity, I recommend to add this information here since readers may have similar doubts as I did.

We thank the reviewer for this remark and have decided to add a clarifying statement at lines 285-286.

Lines 364-370: I highly appreciate that the authors have done this additional analysis to reflect the model-performance skills based on a subset of the data. Yet, I do not agree with the conclusion that most of the observed variance before 2018 originates from stochasticity which also contrasts some of the statements in the results section (e.g. lines 350 and 373). Inspection of Fig. S 3.2.1.2 clearly shows, that the years 2003, 2012, and 2015 were characterized by a higher drought index which is mirrored in increased mortality rates in both species. Thus, the peaks of mortality rates prior to 2018 do not originate from stochasticity but an increasing DI which also becomes evident if inspecting Fig. S 3.3.2. As I see it, this mirrors the threshold-like behavior of the mortality model which – as the authors conclude themselves – results in an overestimation of mortality for beech in general and for spruce in the years prior to 2014. Moreover, I would not call the increase of r^2 an artifact but rather a mathematical feature of the calculation of correlations. The mortality rates following 2017 increase the error squares by an order of magnitude and if this happens in both simulated and observed mortality, r^2 (and MAE) will increase substantially. Since your best AWC-scenario selection Is based on r^2 (and MAE), this is a crucial aspect to consider. Based on Fig. S 3.3.1, it seems that if the scenario selection had been based on the years prior to 2018, different scenarios were selected. The question then arises, how these scenarios performed in the years after 2017. I am not asking the authors to change the scenario selection for the main text at this point, but I strongly recommend to reflect this behavior more clearly in the results and discussion (see also my main comment above). Ideally, the authors would show two supplementary display items, resembling Figs. 3 and 4 but for the shorter calibration period 2000-2017. My concern is that if at any point the aim is to use such kind of mortality models for projections into the future, I would rather rely on AWC-scenarios whose r^2 also performs acceptable under less extreme conditions than 2018. Otherwise, the selected AWC-scenarios might result in too extreme mortality scenarios. Since a model validation based on a subset of data is a common standard, I see the reflection of uncertainty of model performance metrics (r^2 and MAE) as a mandatory point to reflect in terms of clarity.

We thank the reviewer for his comment. As motivated in the answer above, we modified our statement in both the Results (lines 371-378 for beech; lines 451–455 for spruce) and in the Discussion (lines 493–497 and 592–595) sections.

Line 378: to ease the reading I suggest to add ‘combined’ before ‘long- and short’ or revise into: the interplay of effects from drought-effects acting on longer and shorter time-scales.

We have accepted the suggestion and modified the text accordingly at line 382.

Lines 438-444: see my comment on lines 364-370. For spruce, it is also interesting to note, that the increasing observed mortality after 2012 is not captured by the model. Based on the higher complexity of the AWC-scenario performance of Norway spruce I am even more curious to know what would happen if selecting the models based on performance metrics derived from the period 2000-2017. Again, I am not asking the authors to change the key display items of the results section but I would highly appreciate a more comprehensive model falsification (see my suggestions above).

We provide a detailed answer on spruce behavior for the falsification experiment in the detailed answer above.

Lines 450-458: I highly appreciate the inclusion of this sensitivity analysis. The new results totally make sense wherefore they strengthen your model implementation.

We thank the reviewer for this feedback.

Line 468: While I generally agree with this statement, I want to highlight that roughly one third of the mortality observed in 2019 stemming from PIC arises from inciting factors. Also, I am puzzled about the fractions presented in Fig. S 3.3.2. I assume, that total mortality fraction adds up to 100% in each year. So, is the remaining part (which is not shown but adds up to 100%) stochastic mortality? At least I don't fully get how this SI-figure goes together with the main display items which show different temporal dynamics of the mortality rates. Could the authors please elaborate on this? Maybe this also relates to your interpretation of stochasticity dominating the mortality before 2018 so getting this right is quite essential.

We thank the reviewer for pointing this out. We clarify that Figure S 3.3.2 does not represent the full annual mortality ($T_{mortality}$) summing to 100%. Instead, it shows the fraction $\bar{T}_{PIC}$(% of dead trees per year) with DBH ≥ 40 cm, flagged for predisposing and inciting stress factors in the PIC framework. Hence, it only includes the subset of dead trees for which a mechanistic cause could be assigned (slow growth, drought memory, or inciting drought stress).

The remaining proportion of dead trees not accounted for in this figure stems from (i) other size classes, and (ii) trees that died without meeting PIC thresholds, i.e., considered as background/stochastic mortality. This explains why the total bars do not sum to 100%.

We have now clarified this in the figure caption (lines 112-114) to avoid misinterpretation. This distinction also aligns with our interpretation that before 2018, much of the mortality arose from stochastic processes, while the post-2018 mortality increasingly involved trees meeting specific drought-related stress thresholds.

Lines 479-486: I appreciate this critical reflection of the newly added subsample-analysis. Yet, I again want to stress that the patterns before 2018 do not only arise from stochasticity but appear to be triggered by drought events (e.g. 2012, 2015). If implementing my suggestion of additional supplementary display items resembling Figs. 3 and 4 but for the period 2000-2017, the authors could elaborate the discussion in this section to reflect how the period selection affects the AWC-scenario selection. This would also guide future research that mimics the presented approach and address aspects of uncertainty related to a mortality projections based on the AWC-scenarios. I understand that this is not the main point of the manuscript, but it will help to correctly guide related future work.

We appreciate once again the reviewer's comment. We have integrated these considerations in the Discussion section, lines 493–497 and 592–595, respectively.

Lines 572-576: I highly appreciate this paragraph to foster cautious reproduction of the AWC-scenario approach. Maybe the points I suggested for lines 479-486 could instead be added here.

We have added these considerations in the Discussion section lines 493–497 and 592–595, respectively.

Line 586: please check this sentence: ‘but it does induce longer die-off once water availability improves’. This does not make sense in combination with the following sentence. Or is it simply the word ‘die-off’ which is ambiguous in context of mortality discussions? That is, the stress dies off and not the trees? Please clarify and revise.

We thank the reviewer for pointing out the ambiguity in this sentence. Our intention was to describe the persistence of physiological stress symptoms even after drought conditions subside, not a continued increase in tree mortality. We have revised the sentence accordingly to avoid confusion with the term "die-off" at lines 590-591.