Reviewer response

Dear authors,
We have received two follow up assessments from the reviewers. One reviewer has remaining comments that should be considered. I've read the revised manuscript myself and can agree especially with both of the points 1 and 2 that the reviewer raises. I did recognize the three parameterization approaches myself but did not follow where or how that analysis influenced the results.
A final minor point. In line 21 of the tracked changes document, explained variance is mentioned but explained variance of what specifically? Explained variance, within a time period, across years, or across simulations?
We look forward to receiving your revisions.
Sincerely,
Andrew Feldman

We thank the editor for the careful reading and for giving us the opportunity to revise our manuscript. We clarified the term in the abstract (lines 21-22) and Methods (lines 303–307). For each response variable and time period, we fitted one GAM to annual values pooled across all scenario and GCM runs. "Variance explained" refers to the share of each driver or interaction in the summed penalized F statistic × degrees of freedom of that model. The decomposition therefore partitions variation among years and simulation runs within a period, not variation across periods. We replaced "variance explained" with "relative importance" in the abstract.

Regarding the three parameterization approaches, please see our response to Reviewer 2, points 1 and 2.

Reviewer 1

I am satisfied with the authors' responses to the initial review. The manuscript has undergone meaningful revisions that address all major concerns and strengthen its rigor and accessibility.

The validation metrics are now appropriately presented in the main text (Table 1, seasonal breakdown in Table A1, section 3 text), enabling readers to transparently evaluate model performance and assess confidence in projections. This was the primary concern and has been resolved. Model justification has been clarified,

particularly regarding ED2.2-hydraulics suitability. The expanded discussion in Section 4.4 explicitly acknowledges model limitations, notably spring GPP underestimation and summer ET discrepancies, their physiological origins, and implications for projection uncertainty, which substantially improves credibility.

Methodological thresholds are now supported with literature and clear reasoning, post-dieback dynamics are explained mechanistically with reference to empirical thinning studies, and SSP5-8.5 is appropriately framed as an upper-bound stress test.

The scientific importance of this work remains strong: this study now provides a more rigorous benchmark for dryland forest projections under climate change, addressing a critical gap in the literature.

We thank the reviewer for the careful reading of our revised manuscript and the positive assessment as well as for the constructive comments that improved the manuscript.

Reviewer 2

I thank the authors for the careful and thorough response. Most of my comments have been addressed, and the added figures and the revisions to the manuscript helped a lot. I am happy to see the work move forward.

We thank the reviewer for the constructive assessment and for recognising the challenges of applying ED2.2 in drylands. We addressed all three points and expanded Table A2 (now Table S2). Again, your comments helped to improve the clarity and reproducibility of our manuscript and we appreciate your thorough review. Please note that we renamed the supplementary material according to the Biogeosciences guidelines so they have slightly different names in our answers below.

I would, however, like to leave a few notes for the authors' consideration. Calibration and validation remain the main open area. I appreciate that ED2 is a complex model and that this is one of very few applications to an arid system, so I am comfortable passing some of this to the reader with the limitations clearly stated. Still, a few items would be worth reporting if they are readily available:

1. Section 2.4 states that parameters without direct measurements were "identified as influential through sensitivity analysis," but I could not find the sensitivity analysis in the

manuscript or supplement, neither the method used nor the results. The choice of method itself is the subject of a considerable literature (e.g. Sobol vs Morris), and different approaches can lead to quite different outcomes, so even a brief statement of what was done would help.

We agree that the sensitivity analysis was not sufficiently described. Due to computational limitations we used local sensitivity analysis for each parameter instead of a global sensitivity approach. A global method such as Morris or Sobol would have been preferable, but it requires thousands of model runs, which was not feasible. We therefore screened parameter sensitivity manually one at a time on the ED2.2 version without plant hydraulics, varying each vegetation parameter within its literature range while holding all others at reference values. Appendix S1 now describes the procedure, and Section 4.4 lists the absence of a global sensitivity analysis as a limitation (lines 621–624).

2. On the same note and in the same section, three parameterization approaches are described: (i) direct substitution with measured values, (ii) calibration within literature-informed ranges, and (iii) calibration of parameters identified by sensitivity analysis. It is not clear in the Results, or in the SI, which parameters fell into which category. The SI states that "calibration included iterative adjustment of physiological traits," but the nature of that iteration is not specified: was this a numerical optimization (e.g. a gradient-based method), or manual adjustment? It would be informative to see how the three approaches lead to different outcomes. The SI also mentions that optimization focused on "the three best-performing model versions," and I am not sure how to reconcile that statement with the three approaches described above, as they appear to be different things.

We thank the reviewer for this comment and agree that we could have been clearer on where each approach fed into which parameters. Table S2 (formerly A2) now assigns each parameter to approach (i), (ii) or (iii) and gives its value, source and justification.
Calibration was manual for the same computational limitation mentioned above. An exploratory Bayesian calibration did not converge within the 5 day wall time of our cluster system and served only to guide the direction of parameter adjustment. We tried many approaches to make it work within the time limit but unfortunately it was unsuccessful. In the manual calibration, each candidate parameter set was evaluated with RMSE, R^2 and bias against monthly water and carbon fluxes, soil moisture and LAI. The "three best performing model versions" were the three candidate sets with the lowest but similar RMSE, from which we selected the final set by agreement in seasonal pattern and magnitude. The three parameterization approaches describe how individual

parameter values were obtained; the three model versions are complete parameter sets compared at the final selection step. Appendix A1 now separates them more clearly.

3. Table A2 also looks incomplete. For example, soil hydraulic properties such as water retention and conductivity are mentioned in the manuscript (lines 130-133) with reference to Table A2, but do not appear there.

We thank the reviewer for spotting this. We now added to Table S2 (former A2) all the soil configuration and hydraulic properties (layer depths, lower boundary condition, retention exponent, saturated hydraulic conductivity, water content at saturation, field capacity and wilting point), the plant hydraulic traits, and the rooting depth parameters (see also our response to point 2). Section 2.2 now states that water retention and conductivity followed the Campbell–Mualem model with Cosby et al. (1984) pedotransfer functions (lines 134–135).

Again, I fully understand the limitations of applying this model in drylands, and I think the authors have done tremendous work to push this forward. Most of my points ask for clarification on calibration and validation that should not require new model runs; revising the text would suffice at this stage. I do, however, encourage a comprehensive calibration and validation exercise in future work, to shed more light on the key parameters and the model structure challenges specific to these ecosystems. As the authors note, the study site is very well suited to this, with decades of data and active management.

Thanks

We agree and thank the reviewer for the encouragement. A formal calibration and global sensitivity analysis, including the hydraulic parameters, is a priority for our future work at Yatir, given its long-term data availability. We are already exploring options for using a different cluster system or migrating our workflow to a local machine that is not subject to run time limitations.
We again thank the reviewer for their valuable comments and their insight on our study that significantly improved the quality of this paper.