Editor response letter

Dear Dr. Feldman,

We are pleased to submit a revised version of our manuscript "CO_2 Fertilization to Climate Limitation: Shifting Drivers in a Dryland Forest". The reviewer comments were very helpful in improving the clarity and interpretability of the study, particularly regarding the confidence in the model predictions and the presentation of model capabilities and limitations using ED2.2-hydraulics. In response, we substantially expanded the Calibration section to provide a more detailed evaluation of the precipitation- and season specific responses of carbon fluxes. We also revised the Discussion to more clearly highlight the implications for dryland forest management and to provide a more balanced discussion of the model's limitations.

We believe these revisions have strengthened the manuscript considerably and address the key concerns raised by both reviewers. We thank you and the reviewers for the constructive feedback and look forward to your consideration of the revised manuscript. Author comments are shown in italics; manuscript additions appear in blue and L refers to the line numbers in the revised manuscript.

Sincerely,

Dr. Katja Irob, on behalf of all authors

Reviewer response

Reviewer 1

This study projects future drought impacts on the Yatir Forest afforestation using the ED2 ecosystem model under high-CO_2 and warming scenarios. Understanding whether dryland afforestation can maintain productivity and carbon sequestration under intensifying drought is critical for climate adaptation policy in water-limited regions. However, I raise a main concern that the model's calibration/validation may be inadequate—specifically, that GPP and ET dynamics are not well captured even in the improved hydraulic version—which fundamentally undermines confidence in the projections.

Here are the main points to address:

1. The model was calibrated against 20 years of FLUXNET data, including ET and carbon fluxes (Section A2), but critically, the validation metrics (R^2, RMSE, bias, correlation for GPP, ET, soil moisture, LAI) are not reported in the main text or as a table**.** Readers cannot assess whether ET is well-captured (R^2 = 0.8?) or mediocre (R^2 = 0.5?), or whether GPP peaks are matched or missed. These numbers appear only in the supplementary repository, requiring readers to download and run code to assess model performance. This is problematic because readers cannot evaluate whether key fluxes are well-captured or whether discrepancies might propagate into projections.

2. The authors acknowledge limitations in GPP representation (Figures 2 and S2 show seasonal peaks that are missed, particularly in spring and early summer). However, the quantitative extent of these mismatches is unclear in the main text. Line 278 reports that 'model–data agreement was strongest for gross primary productivity (GPP; 89.3%, Fig. 2c),' but what do the seasonal error metrics show? For example, if spring GPP RMSE is ±25% while annual RMSE is ±10%, this distinction matters: spring is when the forest builds up carbohydrate reserves critical for drought survival. Without explicit RMSE, R^2, and bias reported by season in the main text, readers cannot assess whether the model's representation of pre-drought productivity is sufficient to trust projections of forest resilience. Given that GPP mismatches propagate into all downstream carbon projections (NEP, NPP, mortality timing), quantifying these errors explicitly would strengthen confidence in the study's conclusions—or appropriately flag where uncertainty is largest. Similar considerations apply to ET validation.

We thank the reviewer for these important points. We initially left some calibration metrics for the supporting material because we thought this content might be too technical for the readers, but we agree that standard validation metrics must be accessible in the main text. We have added Table 1 to the Methods section reporting R^2, RMSE, and bias for all evaluated variables, and Table A1 (previously Table S1) in the Supplementary Information reporting seasonal RMSE, absolute bias, and relative bias for GPP and ET.

The calibrated model shows strong agreement across most variables (GPP: R^2 = 0.86, RMSE = 0.70 g C m^{-2} day^{-1}; NEP: R^2 = 0.79; soil moisture: R^2 = 0.89; LAI: R^2 = 0.92 over the growing season November–May), which we now also present in more detail in the first section of the Results. ET showed moderate overall agreement (R^2 = 0.70, r = 0.84), indicating that the seasonal shape is well captured while absolute magnitude is less well reproduced. Seasonal decomposition (Table A1, previously Table S1) reveals that summer ET is the primary source of discrepancy (bias: −11.9 mm $month^{-1}$, −66% relative to observed mean of ~17 mm $month^{-1}$). As discussed in the revised manuscript (Section 4.4), this reflects the model's effective shutdown of transpiration under combined soil moisture and VPD stress, missing residual fluxes from brief morning stomatal opening, stem water storage drawdown, and cuticular water loss documented at Yatir (Klein et al., 2016; Preisler et al., 2021, 2022). Since summer carbon fluxes are near

zero and soil moisture bias remains small (+2.8%), this limitation is unlikely to alter projected carbon trajectories but may cause slight underestimation of cumulative hydraulic stress through summer.

Spring GPP underestimation (bias: −1.15 g C m^{-2} day^{-1}, −38%; Table A1, previously Table S1) is the most consequential seasonal error for the study's projections, as spring is when carbon uptake is highest. But we established that this bias is mainly caused by wetter years (Figure A4). As acknowledged in the Discussion, this reflects fixed physiological parameters that cannot reproduce the nonlinear relaxation of all environmental constraints simultaneously in late winter and early spring, a limitation also reported in other ED2.2 applications to dryland systems (Dashti et al., 2021; Pandit et al., 2018). This results in a potential underestimation of early-century CO_2 fertilization benefits, which we consider a conservative bias with respect to the study's main conclusions about forest vulnerability.

We have also added a more elaborated discussion of the fit and limitations of ED2.2-hydraulics for the dryland forest Yatir.

3. ED2 was developed for tropical and temperate forests and is rarely applied to semi-arid ecosystems. While the authors added a hydraulic module to improve water-stress representation, the model's core structure (patch-level demography, implicit hydrodynamic coupling) may not be optimal for extreme drought environments where water availability is the primary limiting factor. The calibration difficulties (see above) raise the question: would a simpler, water-focused model (e.g., dynamic global vegetation model with explicit soil-water balance) or a site-optimized model be more appropriate? The authors should justify why ED2 is suitable despite these constraints, or acknowledge this as a limitation.

The reviewer's concern about implicit hydrodynamic coupling accurately describes standard ED2, but are not that relevant for the version used here. ED2.2 with hydraulics (Xu et al., 2016) explicitly resolves xylem water potential gradients, stomatal regulation tied to hydraulic status, and species-specific vulnerability curves, replacing the empirical water stress formulation with mechanistic plant hydraulics suited to extreme drought.

ED2.2 with hydraulics is well suited to the research questions for three reasons. First, the central findings, including the lag between functional and structural decline, cohort-level dieback driven by unsustainable maintenance respiration costs, and demographic feedbacks between stand structure and carbon balance, emerge from ED2.2's size-structured patch demography. A simpler water-focused DGVM cannot simulate these processes regardless of flux tower agreement. Second, many simpler dryland models operate on annual timesteps and cannot resolve the seasonal dynamics central to this study; ED2.2's sub-daily to monthly temporal resolution is required for capturing the spring productivity window, summer hydraulic stress accumulation, and their interaction with CO_2 fertilization across decades. Third, simultaneous calibration against carbon fluxes, water fluxes, vegetation structure, and radial growth at this arid site (aridity index ~0.18) achieved strong agreement across all components (R^2 range 0.70

to 0.92; Table 1), a validation attempt rarely done in dryland modeling for a model as complex as ED.

Though rarely applied to dryland systems, ED2.2 has proven reliable for water-carbon dynamics in water-limited environments (Chang et al., 2025; Xu et al., 2016; Dashti et al., 2021; Pandit et al., 2019). A dedicated dryland version incorporating photodegradation, hydraulic redistribution, and dynamic rooting remains absent from the literature and is one of our priorities for future development. This study represents a step toward establishing ED2.2-dryland as a viable demographic modeling framework for dryland forests under climate change. We have added brief clarifying statements to the Methods and Discussion (Section 4.4).

L 125–128: Unlike simpler water-focused models that operate on annual timesteps, ED2.2-hydraulics resolves sub-daily flux dynamics and size-structured demography, both of which are required to simulate the cohort-level mortality and lag between functional and structural responses that are central to our research questions.

Other comments:

1. Several decision thresholds need supporting references and justification. The 15-year rolling window for defining 'significant decline' (lines 237–239) is presented without explanation—why 15 years rather than 10 or 20? Does literature on forest productivity dynamics support this window, or does sensitivity analysis show robustness to alternative windows? The 0.5% threshold for NEP collapse similarly needs supporting references. Additionally, the time-lag assumptions discussed on line 242 should cite physiological or phenological literature documenting these delays in Mediterranean pines or similar ecosystems.

We thank the reviewer for this comment. We address each point in turn.

The 15-year rolling window was chosen to balance two competing requirements: filtering the high interannual variability characteristic of dryland systems, where single drought years can temporarily suppress productivity without indicating sustained decline, while remaining short enough to detect meaningful trends within the projection period. Windows of 10 to 20 years have been used in comparable forest productivity trend analyses (Anderegg et al., 2020; Reichstein et al., 2013), and 15 years represents a reasonable midpoint that satisfies both requirements simultaneously and resulting in similar outcomes. AGB growth stagnation thresholds reflect the biologically motivated criterion that net growth has effectively ceased, where new biomass production no longer exceeds turnover and mortality losses. We have added a brief clarifying sentence to the methods to make this distinction explicit.

L264–270: Slope thresholds were informed by the magnitude of interannual variability observed in the historical simulation period (2000–2020), representing the minimum rate of sustained change distinguishable from natural drought-driven fluctuations. A 15-year window was chosen to filter the high interannual variability characteristic of dryland systems, where single drought years can suppress productivity without indicating sustained decline, while remaining short enough to detect meaningful trends within the projection period; windows of this length are

consistent with multi-year lag analyses in comparable forest productivity studies (Anderegg et al., 2020; Reichstein et al., 2013.

L270–273: For cumulative structural variables (total AGB, mean DBH), we detected growth stagnation when 15- year rolling window relative growth rates fell below 0.5% yr^{-1} for AGB and 0.2% yr^{-1} for DBH, reflecting the point at which new biomass production no longer exceeds turnover and mortality losses, effectively indicating the cessation of net structural growth.

For the time-lag assumptions, the physiological basis for multi-year delayed structural responses to functional decline is well documented in the literature cited in the manuscript: stomatal closure and reduced carbon assimilation precede growth reductions by months to years due to carbon storage buffering and structural inertia (Anderegg et al., 2019; Hartmann et al., 2022), and lagged mortality following hydraulic stress has been observed in P. halepensis specifically (Klein et al., 2016; Tatarinov et al., 2016). We have added these citations explicitly at lines 277 to 278.

L276–279: The physiological basis for such lags is well established: stomatal closure and reduced carbon assimilation precede structural growth reductions by months to years due to the buffering effects of carbon storage in the trees and structural inertia (Anderegg et al., 2019; Hartmann et al., 2022), and lagged mortality following hydraulic stress has been documented in P. halepensis at Yatir specifically (Klein et al., 2016; Tatarinov et al., 2016.

2. The choice of SSP5-8.5 (very high emissions, weak mitigation) also warrants discussion. This scenario is now considered less likely than moderate pathways (SSP3-7.0, SSP2-4.5) given recent climate policy. While upper-bound risk scenarios are valuable, how sensitive are key findings (dieback onset, NEP collapse timing) to scenario choice? A brief comparison across scenarios—or explicit framing of SSP5-8.5 as an upper-bound risk rather than a most-likely forecast—would help readers contextualize the conclusions.

We agree that SSP5-8.5 is best understood as an upper-bound stress test and should not be regarded as a central projection, consistent with IPCC AR6 guidance (IPCC, 2021). However, SSP5-8.5 in terms of physical processes remains valid and can be useful to test uncertainty and limits. We note that the manuscript already describes SSP5-8.5 as representing "the upper end of plausible radiative forcing by the end of the century" in the methods section. SSP5-8.5 was included precisely to characterize the risk envelope under high-end forcing, not as a most-likely forecast. SSP3-7.0 forms the basis for our primary discussion of decline timing and driver transitions. We further note that at the time of submission, SSP5-8.5 remained standard practice in the climate impacts literature and was only formally removed as a priority scenario by the CMIP7 ScenarioMIP framework in April 2026, after our manuscript was submitted. The manuscript compares both scenarios explicitly for all key findings. We have nonetheless added a sentence to the methods making the upper-bound framing more prominent, as the reviewer

suggests. Regarding the sensitivity of key findings to scenario choice, the pattern of all central results holds under both scenarios, mostly expressed in different magnitudes.

L179–181: SSP5-8.5 was included as an upper-bound stress test to characterize the risk envelope under high-end forcing, consistent with its intended use in IPCC AR6 (IPCC, 2021), while SSP3-7.0 represents a more policy-relevant intermediate pathway for the primary assessment of decline timing and driver transitions.

3. Line 338 describes 'spikes' in variables following dieback events, but the mechanism is unclear. Do remaining trees compensate through increased growth? Is this an artifact of the model's mortality algorithm releasing light and water to survivors? What is the magnitude, and is it biologically realistic? This needs explicit explanation because it affects interpretability of post-disturbance dynamics.

Yes, the spikes in LAI and WUE following dieback events reflect resource release to surviving cohorts after patch-level mortality, a real process represented mechanistically in ED2.2's demographic structure. When mortality eliminates the largest trees, surviving cohorts experience reduced competition for soil water and increased light availability, driving a transient recovery in leaf area and carbon assimilation. The WUE spike is partly a ratio effect: as stand density drops, transpiration falls faster than GPP in surviving individuals under improved water availability, temporarily inflating WUE. This post-disturbance growth release is biologically realistic and consistent with experimental thinning results in Israeli dryland P. halepensis plantations, where density reduction improved growth and reduced mortality, with effects scaling with thinning intensity (Calev et al., 2016). These findings have since informed thinning as an endorsed management intervention in Israeli dryland afforestation. We have added a sentence at the point of first description of the spikes clarifying the mechanism and noting that the recovery is transient, as surviving trees face the same long-term climate trajectory without regeneration.

L565–568: The LAI recovery following mortality events in our simulations suggests that density reduction can temporarily restore canopy function, consistent with experimental thinning results in dryland P. halepensis plantations showing improved growth, reduced drought sensitivity, and lower mortality (Calev et al., 2016), findings that have since informed thinning as an endorsed forest management intervention in Israeli dryland afforestation.

4. Line 346 states that NEP decline onset is later under SSP5-8.5 than under milder scenarios (e.g., SSP3-7.0). This is counterintuitive: we would expect the most severe warming scenario to show earlier and more severe decline, not later. The authors note "larger variability between simulations" under SSP5-8.5, but this does not explain why the mean decline onset shifts later under the warmest scenario. Is this due to fewer ensemble members converging under SSP5-8.5, creating a statistical artifact? Are there nonlinear dynamics (e.g., early hydraulic failure followed by brief recovery before

collapse) that delay mean onset? Is this an error in temporal alignment or scenario attribution?

We thank the reviewer for raising good questions, we address each of the reviewer's questions in turn.

The apparently counterintuitive later NEP decline was detected in only 3 of 5 SSP5-8.5 models; the mean of year 2053 ± 16 is computed from these three only, as is standard when not all ensemble members trigger the threshold. Two models showed no sustained NEP decline within the century, reported explicitly as non-detections. Under SSP3-7.0 all five models start declining in the year 2037 ± 0, producing a more robust mean.

Higher CO_2 concentrations under SSP5-8.5 is an example for nonlinear dynamics delaying mean onset. Higher atmospheric CO_2 buffered carbon assimilation in surviving individuals under reduced competition, delaying sustained functional decline in some projections despite more severe climate stress, consistent with the driver decomposition (Fig. 5, previously Fig. 4). Additionally, gfdl-esm4's late detection (2071) reflects abrupt stand-level collapse being captured by the rolling window only after the trajectory became catastrophic, pulling the mean later.

No, it is not an error in temporal alignment or scenario attribution. We have verified the scenario attribution and temporal alignment in Tables B1a and B2b (previously Table S2); the pattern is robust.

We added the following to Section 4.1:

L509–515: Under SSP5-8.5, the stronger CO_2 forcing produced a counterintuitive pattern: mean NEP decline onset was later (2053 ± 16) than under SSP3-7.0 (2037 ± 0), despite more severe climate conditions. This reflects trajectory divergence across models rather than greater forest resilience. Two of five models showed no sustained NEP decline within the century, and among the three that did, gfdl-esm4's late detection (2071) reflects abrupt stand-level collapse captured by the rolling window only after the trajectory became catastrophic. In the remaining models, the stronger CO_2 fertilization under SSP5-8.5 buffered carbon assimilation in survivors under reduced competition, temporarily delaying sustained functional decline, aligning with the driver decomposition showing sustained CO_2 contributions to NEP into late century under SSP5-8.5 (Fig. 5).

5. The paper provides valuable analysis of which drivers (CO_2, temperature, water stress) explain variance among simulations (Figure 4). However, equally important would be understanding what drives the mean projected change and how important each driver is to the central estimate. A complementary analysis (e.g., a figure similar to Figure 4 but showing driver contributions to mean NEP and GPP trajectories) would clarify whether the ensemble mean decline is primarily driven by water stress, temperature, or declining CO_2 benefit would be really appreciated.

We thank the reviewer for this suggestion. To produce a mean driver attribution, we attempted three approaches: fitting the GAM directly to the ensemble mean trajectory per scenario; extracting partial effects from the existing GAMs averaged across GCM runs; and fitting a single GAM across all years with GCM as a random effect. All three approaches failed to reliably attribute water stress contributions to NEP, because water stress operates primarily through nonlinear interactions with CO_2 and temperature rather than as an additive main effect. With only five GCMs, fitting interaction terms on the mean trajectory lacks sufficient statistical power.

The cross-simulation variance decomposition in Figure 5 (previously Figure 4) captures how differences in climate trajectories across GCMs, within a shared CO_2 forcing pathway, shape carbon flux outcomes. The consistent dominance of water stress and the transient CO_2 buffering role across models therefore reflect the forest's mean sensitivity to these drivers across the plausible range of future climate conditions, making Figure 5 informative about the drivers of the central estimate as well as the spread.

We added the following to the Discussion section 4.1.:

L474–477: Attempts to decompose drivers of the ensemble mean NEP trajectory directly confirmed that water stress influences NEP primarily through nonlinear interactions with CO_2 and temperature rather than as an independent additive effect, consistent with the compound driver dynamics shown in Figure 5 and precluding clean mean attribution with the available ensemble size.

> Suggestion: The paper analyzes vulnerability—when and how the forest declines—providing important information for understanding forest resilience. An analysis framing these projections in terms of “net carbon benefit” would strengthen the manuscript and enhance its value to policymakers. Over the full projection period, what is the cumulative carbon stored in the Yatir stand under each scenario (Mg C/ha or Tg C total)? If collapse occurs in 2060 under SSP5-8.5, is the cumulative carbon benefit positive or negative relative to the cost of establishment and management? Even a straightforward calculation (e.g., 'SSP3-7.0 yields ~250 Mg C/ha by 2100; SSP5-8.5 yields ~150 Mg C/ha before collapse') would frame results in terms of afforestation viability and help land managers and policymakers weigh uncertainty. This would transform the analysis from a purely vulnerability assessment into decision-relevant information about whether this afforestation is 'worth it' under climate scenarios

We thank the reviewer for this suggestion. Despite periodic source years, the forest accumulated net carbon under both future scenarios, with CO_2 fertilization driving cumulative sequestration despite intensifying drought stress: 106 Mg C ha^{-1} under SSP3-7.0 (range 77 to 139 Mg C ha^{-1}) and 70 Mg C ha^{-1} under SSP5-8.5 (range 62 to 85 Mg C ha^{-1}). We have added these figures to the Discussion to provide the policy-relevant framing the reviewer suggests.

L468–474: Under current climate with constant CO_2 concentration (400 ppm), the dryland forest operated near carbon neutrality, with cumulative NEP of 15 Mg C ha^{-1} over 2015 to 2100 and NEP periodically negative as the dryland forest shifted between operating as either carbon sink

or source. Under elevated CO_2, cumulative sequestration averaged 106 Mg C ha^{-1} under SSP3-7.0 (range 77 to 139 Mg C ha^{-1}) and 70 Mg C ha^{-1} under SSP5-8.5 (range 62 to 85 Mg C ha^{-1}), reflecting the dominant role of CO_2 fertilization in driving net carbon gain over the century, despite intensifying drought stress. Models projecting stand collapse under SSP5-8.5 accumulated 69 to 74 Mg C ha^{-1} before structural decline, only marginally lower than non-collapsing projections, as most sequestration occurred prior to collapse.

L575–577: Afforestation remains an important climate-mitigation tool: under future scenarios, rising CO_2 drove substantial net carbon accumulation well above the current climate baseline, even in projections with eventual stand collapse, as most sequestration occurred before structural decline.

Final comment:

> This study addresses an important question regarding the viability of dryland afforestation under climate change. The analysis is interesting and novel for this ecosystem type, and projecting potential dieback under intensifying drought is a valuable contribution to climate adaptation research. However, the solidity of the results needs to be better established. The validation metrics—R^2, RMSE, bias, and seasonal error estimates—are currently relegated to the supplement repository (not suppl. info), making it difficult for readers to assess whether model discrepancies are acceptable for the study's key conclusions about forest vulnerability and carbon dynamics. These metrics must be brought into the main text so that the credibility of the projections can be properly evaluated. With the revisions addressing model validation and result interpretation, the work could provide an interesting contribution for understanding afforestation resilience in dryland regions.

We thank the reviewer for the constructive overall assessment. The validation metrics have been added to the main text as Table 1 (R^2, RMSE, and bias for all evaluated variables) and Table A1 (seasonal breakdown for GPP and ET), directly addressing the primary concern. The additional revisions addressing model justification, decision thresholds, scenario framing, post-dieback dynamics, NEP decline timing, and cumulative carbon benefit are detailed in the responses above. We believe these additions substantially strengthen the manuscript's credibility and policy relevance.

Reviewer 2

> Summary:

> This paper addresses the response of dryland forests to future climate projections under a range of scenarios. To do so, the authors undertook the substantial task of developing a new P. halepensis PFT for the Ecosystem Demography model (ED2). The paper is timely, as our understanding of vegetation dynamics in drylands, and in particular the ability of models to represent them under current and future climate, remains limited. Most DGVMs, including ED2, lack both the parameters and the processes needed to capture dryland vegetation adequately.
>
> That said, I think the paper would benefit from a more in-depth analysis of model performance, as I outline below. Future projections are only informative if we can first agree that the model reproduces current observations well. I hope the comments below are helpful to the authors.

We thank Referee #2 for the thorough and constructive review. The summary accurately captures the central contribution and challenge of the work: developing and validating a new P. halepensis PFT for ED2.2 in an arid system where most DGVMs lack adequate parameterization. We fully agree that the credibility of future projections rests on demonstrating adequate model performance under current conditions, and the revised manuscript substantially strengthens the validation in response to both referees' comments.

The model used here, ED2.2 with hydraulics (Xu et al., 2016), goes beyond standard ED2 by explicitly resolving xylem water potential gradients, stomatal regulation tied to hydraulic status, and species-specific vulnerability curves, replacing the empirical water stress formulation that the reviewer rightly identifies as inadequate for dryland systems. The new P. halepensis PFT was calibrated simultaneously against carbon fluxes, water fluxes, vegetation structure, and radial growth, achieving R^2 of 0.70 to 0.92 across all components, a validation approach rarely attempted in dryland modeling.

The revisions address all major and minor points raised: standard validation metrics are now reported in Table 1 and Table A1 with seasonal breakdown; interannual validation for dry, typical, and wet years appears as Figure A4 in the supplementary; the LAI figure has been corrected; specific parameter constraints driving the spring GPP underestimation are identified; the dieback mechanism is demonstrated in a new supplementary figure; and an early warning signal figure has been added to Section 3.3.

Major comments:

> SSP5-8.5 is one of the most extreme scenarios. Recent work suggests such high-forcing pathways are increasingly viewed as unlikely, and CMIP7 has retired them from its core set (https://gmd.copernicus.org/articles/19/2627/2026/). I am not asking the authors to redo the analysis, but a strong justification for why SSP5-8.5 remains informative in this context should be provided.

As also noted by Referee #1, SSP5-8.5 was standard practice in the climate impacts literature at the time of submission and was only formally retired from the CMIP7 priority set in April 2026, after manuscript submission. Besides, in terms of physical processes it remains a valid scenario and can be useful to test uncertainty and limits. Van Vuuren et al. (2026) note that the new CMIP7 High scenario is expected to produce forcings below SSP5-8.5, confirming it continues to represent the upper bound of plausible radiative forcing, exactly its role here. Dryland afforestation at the margins of viability is where upper bound stress testing is most policy-relevant, and SSP3-7.0, which more closely tracks current policy trajectories, is included throughout as the basis for primary assessment of decline timing and driver transitions. The Methods section has been updated to make the upper bound framing explicit and to cite Van Vuuren et al. (2026).

L179–181: SSP5-8.5 was included as an upper-bound stress test to characterize the risk envelope under high-end forcing, consistent with its intended use in IPCC AR6 (IPCC, 2021), while SSP3-7.0 represents a more policy-relevant intermediate pathway for the primary assessment of decline timing and driver transitions in CMIP7 (Van Vuuren et al., 2026).

> In the Results, it is unclear why the model was averaged over 20 years (e.g., Figure 2 and A2). Dryland forests are highly sensitive to environmental variability such as precipitation, as the authors themselves emphasize. It would be more informative to see whether ED2.2 with the new PFT captures seasonal patterns when they matter most, for example during exceptionally wet or dry years. Additional analysis of temporal dynamics such as phenology, trends, and interannual variability would strengthen the validation. The main text also states that R^2 and RMSE values are provided in SI A2, but I could not locate them there. The reported agreement scores are somewhat misleading. LAI shows 86.6% agreement, yet Figure 2 makes clear that the seasonal dynamics, which are the key feature, are not captured well. GPP has the highest agreement score, but Figure 2 suggests this is driven by low-activity months where GPP is near zero, while during the active season the simulation falls well below observations. A more robust validation is needed before moving to future projections.

We thank the reviewer for these points, which we address together as they concern the same validation framework.

On the 20-year averaging in Figure 2: It is important to highlight that we computed all the validation metrics, R^2, RMSE, and bias, on the full monthly time series across all years (n = ~250 month-year pairs), capturing interannual variability in model performance. Figure 2 shows the climatological mean seasonal cycle for clarity: simply because we thought that plotting all 21 annual time series would produce an unreadable figure given the high interannual variability characteristic of dryland systems.

To address the reviewer's request for interannual performance, we have added Figure A4 to the supplementary A1, showing observed versus simulated GPP and NEP for a dry year, a typical year, and a wet year. Agreement in the dry and typical years is high and essentially unbiased: GPP R^2 = 0.84 and 0.89 with mean bias of +0.13 and −0.12 g C m^{-2} day^{-1}, and NEP R^2 = 0.91 and 0.82 with mean bias of +0.17 and −0.13 g C m^{-2} day^{-1}. The wet year is the only one showing substantial bias (GPP R^2 = 0.73, bias −0.96; NEP R^2 = 0.68, bias −0.45), roughly seven to eight times larger, driven by an observed March to April productivity pulse that exceeds the simulation by about a factor of two. This reflects the nonlinear response to simultaneous relaxation of soil moisture, temperature, and VPD constraints that fixed phenological parameters cannot reproduce, a limitation acknowledged in Section 4.4. Because the projections follow a drying trajectory, the model performs best under precisely the conditions that dominate the simulated future, making the wet-year bias conservative with respect to our conclusions about forest vulnerability.

On the LAI agreement score: We acknowledge an error in Figure 2: the LAI panel showed the model climatology averaged over 21 years plotted against Sprintsin et al. observations covering only 2004 to 2006, creating a misleading visual mismatch in the summer lower distribution. We now corrected the figure to show only the model years corresponding to the observation period (2004 to 2006), which substantially improves visual agreement and better reflects the R^2 of 0.92 reported in Table 1.

On the custom agreement scores: The composite scores (e.g. LAI 86.6%) were a custom weighted metric from our calibration workflow combining pattern correlation and magnitude accuracy. These have been replaced in the main text by standard validation metrics (R^2, RMSE, bias; Table 1), which are more interpretable and directly comparable to other studies. The reviewer's concern that GPP agreement is driven by low-activity months is addressed by the seasonal breakdown in Table A1 (previously Table S1), which shows that summer GPP bias is negligible (−3.7%) while spring GPP bias is 38% confirming the active season underestimation the reviewer identified. The reviewer was right that spring bias is the main driver but by looking specifically at carbon fluxes in a dry, average and wet year we saw that the bias is mainly driven by above average wet years.

On missing metrics in SI A2: We apologise for this confusion. The R^2 and RMSE values were available in a supplementary calibration R Markdown file in the code repository but were not included in the manuscript supplement directly. They are now properly reported in Table 1 (main text) and Table A1 (previously Table S1; seasonal breakdown in the supplement).

Section 3.2.1: dieback events appear near 2080 in some projections. The authors attribute this to CO_2-driven biomass accumulation producing large trees with a negative carbon balance. I am not familiar with the site, but I doubt this is a real ecological signal. It more likely reflects a model artifact, for example from allocation equations that may not be appropriate for this PFT. This warrants careful investigation. The Discussion refers to "maintenance respiration costs of the largest trees eventually exceeded their carbon assimilation," but this is never explicitly demonstrated in the main manuscript. Figure S8 shows time series of the

relevant fluxes but does not demonstrate the proposed mechanism directly. A simple plot of GPP against total respiration over time would show whether respiration ever exceeds GPP, if that is indeed the driver.

We thank the reviewer for this question. The reviewer is right that the maintenance respiration framing was not demonstrated. We have added Figure B7, comparing GPP, total ecosystem respiration, and autotrophic respiration for the two collapsing projections (gfdl-esm4 and mpi-esm1-2-hr) over 2060 to 2099. GPP declined sharply in hot and dry years while respiration declined less, but ecosystem respiration exceeded GPP only in isolated years. The sustained signal is a reduction in net carbon gain before collapse instead of one threshold crossing, so we have replaced the maintenance respiration framing throughout with cumulative hydraulic stress driving abrupt mortality of the largest cohorts.

On whether the collapse reflects a model artifact, we found no indication of one. Cohort-level output confirms that mortality was concentrated in trees above 50 cm DBH, so the reviewer is correct that the largest individuals are involved. These deaths were attributed to the hydraulic failure mortality pathway rather than to carbon starvation, and collapse timing differed between the two projections according to when extreme drought years fell in each model's forcing. An allocation artifact would be expected to trigger at a consistent size irrespective of climate, which is not what we observe. We therefore attribute the collapse to hydraulic failure in large, high water demand individuals under sustained drought stress, with CO_2 driven biomass accumulation acting as a predisposing factor rather than a direct cause. In parallel, we observed in the real forest a similar high mortality event just last year.

L359–364: During subsequent hot and dry years, GPP declined sharply whereas ecosystem respiration decreased less strongly (Figs. B6, B7), leading to a marked reduction in net carbon balance before the onset of widespread mortality and AGB collapse. Cohort-level analysis of the collapse events confirmed that mortality was concentrated in trees with DBH >50 cm, which disappeared entirely within one to two years, while mid-size cohorts (40 to 50 cm DBH) survived and expanded their LAI under reduced competition for light and water.

L522–524: Cumulative drought stress drove abrupt hydraulic failure in the largest cohorts, whose high water demand made them most vulnerable once soil moisture fell below critical thresholds; CO_2 driven biomass accumulation predisposed the stand to this outcome by sustaining growth of large, water-demanding individuals beyond what the climate trajectory could support (Fig. 3a.1, 3a.2, B6, B7).

In the same figure, some simulations show AGB dropping while LAI increases abruptly at the same time (Figure 3a.4). This is not a plausible physiological response and again suggests a model issue. One would also expect a sharp change in structural variables to be mirrored by functional variables such as GPP, but this is not the case.

We thank the reviewer for pressing on this, as the pattern might look implausible at the level of an individual tree. But, it reflects stand-level demographic dynamics rather than a physiological inconsistency. Cohort-level output confirms that mortality at the collapse events was concentrated in the largest size classes (DBH >50 cm), while mid-size cohorts (40 to 50 cm DBH) expanded their canopy once released from suppression by the dominant large trees. Stand-level LAI therefore rose despite the loss of total AGB. The pattern is consistent with thinning experiments in dryland P. halepensis plantations, where density reduction increased LAI and productivity of the remaining trees (Calev et al., 2016). We have expanded the sentence where the LAI spikes are first described, adding the clause after the colon:

L379–381: Some of its projections showed episodic spikes after 2080, coinciding with tree mortality events (Fig. 4a.2): selective mortality of the largest cohorts releases light and water for surviving mid-size trees, temporarily expanding stand-level canopy despite the reduction in total AGB.

The reviewer is right to expect a functional response, and our simulations do show one. Both dieback projections show sharp NEP declines, and Figure B7 shows GPP dropping steeply at collapse in both models. Our original wording obscured this, and we have corrected the manuscript accordingly. After the drop, GPP recovers partially over the subsequent years as surviving mid-size cohorts expand into the released light and water, the same demographic mechanism that produces the transient LAI increase. Structure and function are therefore coupled at the collapse events, and the partial recovery reflects stand reorganization and not continued decline in the surviving trees.

L399–400: Stand-level GPP declined sharply at the collapse events in both projections, followed by partial recovery over the subsequent years (Fig. B7).

> Functional responses (e.g., WUE) show noisy behavior across the century with no consistent pattern across models or scenarios.
> Section 3.4: some results here need more careful examination. For instance, how can CO_2 explain 85% of the variance in DBH but close to 0% for AGB, given that these variables are closely correlated? The early/mid/late categorization also appears somewhat arbitrary: CO_2 explains 83% of the variance in mean DBH in the mid period but drops abruptly to 32% in the late period, which is difficult to reconcile with a smoothly evolving system.

On WUE: the interannual variability in WUE reflects its ratio structure, dividing GPP by ET amplifies variance in both components under dryland conditions where both fluxes respond strongly to interannual precipitation variability. The consistent elevation of WUE under future scenarios relative to current climate, and the late-century spikes associated with post-dieback resource release, are discussed in Sections 3.2 and 3.3. A sentence clarifying the ratio-driven noise has been added to Section 3.2.

L404–406: WUE shows high interannual variability throughout the projection period, reflecting its ratio structure: dividing GPP by ET amplifies variance in both components under dryland conditions where precipitation-driven fluctuations affect both fluxes simultaneously.

On CO_2 explaining DBH but not AGB variance: these variables respond differently to CO_2 fertilization through stand dynamics. CO_2 drives individual tree growth, increasing DBH of surviving trees consistently across GCMs hence the high proportion of explained variance for DBH. Stand-level AGB, however, integrates both individual biomass gain and stem density loss through self-thinning: faster individual growth under elevated CO_2 accelerates competitive suppression and mortality of smaller trees, reducing stem density. The net AGB effect is therefore dampened relative to DBH because biomass gains in surviving trees are partially offset by density-driven biomass losses. CO_2 drives DBH variance strongly precisely because DBH captures individual-level growth while AGB captures the net stand-level outcome of competing processes.

On the abrupt CO_2 variance drop from mid to late century in DBH: the transition from 83% to 32% is not a categorization artifact but reflects the central finding of the study, the shift from CO_2 dominance to climate limitation. In the early and mid periods, all GCMs share the same rising CO_2 trajectory while their climate trajectories diverge modestly, so CO_2 explains most between-model variance. In the late period, climate trajectories diverge substantially across GCMs under sustained warming and drying, and water stress becomes the dominant source of between-model variance. The abruptness of the transition reflects the nonlinear nature of drought stress onset rather than a model artifact. A clarifying sentence has been added to Section 3.4.

L433–436: Stand-level AGB variance is less explained by CO_2 than DBH because AGB integrates both individual biomass gain and density-driven biomass loss through self-thinning. That is, future projections show that faster individual growth under elevated CO_2 accelerates competitive suppression of smaller trees, so biomass gains in surviving individuals are partially offset by mortality-driven density reduction.

L442–444: The decline in the variance explained by CO_2 between mid and late century coincides with increasing between-model variance in climate trajectories, particularly in water availability and temperature stress, which account for increasing variability in the late period.

The Discussion provides limited treatment of model performance itself, which is a notable gap given that this study introduces a new halepensis PFT for ED2.2. Section 4.4 attempts to address this but reads as a mix of literature review and brief acknowledgment of results, and I have mixed feelings about how it is currently structured. A diagnosis of how the identified limitations affect the simulations under this new PFT would be more valuable. For example, the spring GPP underestimation is attributed to "fixed physiological parameters," but which ones? Vcmax? Stomatal conductance? Phenology triggers? Without this, the diagnosis is not actionable for others adapting the model. I would also encourage

the authors to offer a prioritized roadmap for model development rather than an undifferentiated list of missing processes.

We thank the reviewer for this constructive critique. We've tried to improve the section to provide a thorough interpretation of the strengths and caveats of our ED2.2 version for a dryland pine forest. We think this is now an important section of the paper, which might be useful for future studies of dryland forests and climate change impacts in drylands.

First we explained how the spring GPP underestimation reflects two specific parameter constraints. That is, P. halepensis is assigned a strictly evergreen phenology (phenology = 0) with fixed seasonal growth elongation factors (spring_elongf = 1.45, summer_elongf = 0.38, drought_factor = 0.68). These factors are constants calibrated to mean seasonal conditions and cannot adjust in response to antecedent winter soil moisture recharge, the model responds to instantaneous climate forcing only, missing the year-to-year shift in spring onset timing documented in wetter years (Fig. A4).

Second, Vcmax (Vm0 = 21.8 µmol m^{-2} s^{-1}) and the Ball-Berry stomatal slope (stomatal_slope = 10.5) were calibrated to reproduce monthly carbon and water fluxes simultaneously across all seasons. Fixed calibration parameters to a range of conditions constrains peak spring photosynthetic capacity when soil moisture, temperature, and VPD simultaneously reach their seasonal optimum. Future parameterization should prioritize dynamic phenological responses to antecedent soil moisture and targeted Vm0 recalibration against spring flux data specifically, as these most directly affect early-century CO_2 fertilization projections.

Following this important comment we also reorganized our model development priorities, setting three future development tiers in the revised Section 4.4: dynamic phenology and hydraulic redistribution as highest priority given their direct effect on projected functional decline timing; photodegradation and post-rainfall respiration pulses as medium priority given their effect on current-climate carbon balance accuracy; and dynamic nutrient cycling and fire disturbance as longer-term structural developments.

L592–607: The model underestimates this spring peak due to two specific fixed parameter constraints. First, P. halepensis is assigned a strictly evergreen phenology with fixed seasonal growth elongation factors (spring = 1.45, summer = 0.38), which prescribe the relative seasonal distribution of growth independent of preceding climatic conditions. As a result, the spring productivity window cannot advance or extend in response to antecedent winter soil moisture recharge, and the model responds to instantaneous climate forcing only. Second, Vcmax (Vm0 = 21.8 µmol m^{-2} s^{-1}) and the Ball-Berry stomatal slope (stomatal_slope = 10.5) are also fixed parameters calibrated to mean seasonal carbon and water fluxes simultaneously, which constrains peak spring photosynthetic responses under exceptionally favorable conditions. Because the 21-year calibration period was dominated by average and dry years, with only four wetter-than-average years, the optimized parameter set primarily reflects the ecosystem's typical hydroclimatic conditions. Together with the fixed phenology and physiology, this limits the model's ability to reproduce the nonlinear productivity response during exceptionally wet years. Fixed physiological parameters that cannot reproduce such nonlinear responses are a known limitation also reported in other ED2.2 applications to dryland systems (Dashti et al., 2021; Pandit et al., 2018), and is most pronounced in anomalously wet years where deeper soil

moisture recharge advances observed spring onset beyond what fixed seasonal elongation factors can capture (Fig. 3). Since projections focus on a drying trajectory where spring conditions become progressively more limiting, this wet-year bias does not propagate into the primary conclusions about forest vulnerability, but results in a potential underestimation of early-century CO_2 fertilization benefits.

L616–621: Future improvements of the model fall into three tiers. Dynamic phenology responsive to antecedent soil moisture and hydraulic redistribution are highest priority, as they directly affect projected functional decline timing and magnitude. Photodegradation and post-rainfall heterotrophic respiration pulses are medium priority, improving current-climate carbon balance accuracy without affecting the primary vulnerability conclusions. Dynamic nutrient cycling, formal parameter optimization, and explicit fire disturbance represent longer-term structural developments that would extend the model's applicability across dryland ecosystems more broadly.

Minor comments:

> Abstract: define GAM at first use.

We changed the description of this sentence entirely, not mentioning the GAM specifically.

> The manuscript alternates between "forest" and "dryland forest" throughout. This is confusing when processes and consequences are described, especially for readers less familiar with dryland systems, which differ from other forest types in important ways.
> Paragraphs three and four of the Introduction, for instance, describe processes without clarifying whether they apply to dryland forests specifically or to forests in general. If the latter, it reads as overgeneralization. Since dryland forests are the focus, the Introduction should stay in that context as consistently as possible.

We thank the reviewer for this observation. The manuscript has been revised to consistently distinguish between general forest processes, which apply broadly, and processes or claims specific to dryland forests. Bare references to "forest" in the Discussion and Conclusions where Yatir or dryland systems specifically are meant have been replaced with "dryland forest" throughout.

> Are these planted forests actively managed? If so, please state it in the paper.

Yatir Forest is actively managed by The Jewish National Fund (KKL-JNF), with periodic thinning, firebreak maintenance, and access management, but without irrigation or fertilization. A statement to this effect has been added to the site description in the Methods section.

L 99–100: Yatir Forest is actively managed by the Jewish National Fund (KKL-JNF), with periodic thinning interventions, firebreak maintenance, and access management. No irrigation or fertilization is applied. Thinning experiments conducted within the forest have informed density management guidelines for dryland pine plantations in Israel (Calev et al., 2016).

> Lines 174–175: was the "ISIMIP3BASD v2.5 statistical bias adjustment" applied by the authors, or is it an external product used as delivered?

The ISIMIP3BASD v2.5 bias adjustment was applied by the ISIMIP3b team[1] [2] [3] , specifically our co-author M. Büchner, and the corrected outputs were obtained directly from the ISIMIP3b data repository. The sentence has been revised to make this clear.

L190–192: The climate data were obtained as bias-corrected outputs from the ISIMIP3b data repository, where the ISIMIP3BASD v2.5 statistical bias adjustment and downscaling method has been applied by the ISIMIP3b team specifically for the Israeli grid (Lange, 2019).

> Lines 194–198: the thresholds used for the drought severity index need justification. The same applies to the temporal periods (2015–2040, etc.). As written, these choices appear subjective.

The percentile thresholds (20th for precipitation and soil moisture, 80th for VPD) follow established conventions for identifying anomalously dry and hot conditions in drought assessments (Spinoni et al., 2019; Zscheischler et al., 2020), and the 5 mm month $^{-1}$ dry season threshold follows the standard Walter-Lieth climatic classification. The temporal periods align with IPCC AR6 near-term, mid-term, and long-term projection horizons (IPCC, 2021), facilitating comparison with broader climate impact literature. Justifying sentences have been added to the Methods.

L216-219: Percentile thresholds (20th for precipitation and soil moisture, 80th for VPD) follow established conventions for identifying anomalously dry and hot conditions in drought assessments (Spinoni et al., 2019; Zscheischler et al., 2020). The 5 mm month $^{-1}$ threshold for dry season classification follows the standard Walter-Lieth convention for defining climatologically dry months.

L229–230: Temporal periods align with the IPCC AR6 near-term, mid-term, and long-term projection horizons (IPCC, 2021), facilitating comparison with broader climate impact assessments.

> Lines 208–212: a dense block of statistical tests is listed without explanation. A brief description of each test and why it was chosen, either here or in the SI, would help the reader.

A brief justification for each statistical test has been added to the Methods. Kruskal-Wallis tests were chosen as non-parametric alternatives to ANOVA given the small ensemble size and ordinal drought severity scores; chi-square tests were used for categorical dominance patterns;

and Dunn's tests with Bonferroni correction were applied for post hoc pairwise comparisons to control family-wise error rate.

L230–233: Changes in drought severity distributions between these periods were assessed within each scenario using a Kruskal–Wallis test. A non-parametric alternative to one-way ANOVA appropriate for comparing distributions across more than two groups without assuming normality, given the small ensemble size (N = 5 GCMs) and ordinal drought severity scores.

L235–236: Dunn's tests with Bonferroni correction were applied for post hoc pairwise comparisons to control the family-wise error rate across multiple comparisons.

Section 2.6.1: how was the drought decomposition actually done? This is not clearly explained.

The drought decomposition method is described in detail in the Statistical Analysis section (Section 2.6), where the four binary stress indicators, percentile thresholds, and composite scoring approach are defined. Section 2.6.1 has been revised to include a cross-reference to that section at the point where the multicomponent analysis is first mentioned, clarifying that the drought severity index underlying the decomposition is constructed there.

L239–240: Drought severity was quantified using the multicomponent index described in Section 2.6., combining four binary stress indicators (low precipitation, low soil moisture, extended dry season, and high VPD) into a composite score ranging from 0 to 4 for each year.

Lines 237–240: the rationale for the specific threshold values is unclear.

We thank the reviewer for this comment that has also been raised by a second reviewer. The rationale for the slope thresholds and rolling window has been added to the Methods section. Slope thresholds were informed by the magnitude of interannual variability observed in the historical simulation period (2000–2020), representing the minimum rate of sustained change distinguishable from natural drought-driven fluctuations. The 15-year window was chosen to filter high interannual variability characteristic of dryland systems while remaining short enough to detect meaningful trends, consistent with comparable forest productivity studies (Anderegg et al., 2020; Reichstein et al., 2013). Growth stagnation thresholds (0.5% yr^{-1} for AGB, 0.2% yr^{-1} for DBH) reflect the point at which new biomass production no longer exceeds turnover and mortality losses. As also noted by Referee #1, the 0.5% threshold applies to AGB growth stagnation, not NEP collapse.

L265–274: Slope thresholds were informed by the magnitude of interannual variability observed in the historical simulation period (2000–2020), representing the minimum rate of sustained change distinguishable from natural drought-driven fluctuations. A 15-year window was chosen to filter the high interannual variability characteristic of dryland systems, where single drought years can suppress productivity without indicating sustained decline, while remaining short enough to detect meaningful trends within the projection period; windows of this length are consistent with multi-year lag analyses in comparable forest productivity studies (Anderegg et al., 2020; Reichstein et al., 2013). For cumulative structural variables (total AGB, mean DBH),

we detected growth stagnation when 15- year rolling window relative growth rates fell below 0.5% yr^{-1} for AGB and 0.2% yr^{-1} for DBH, reflecting the point at which new biomass production no longer exceeds turnover and mortality losses, effectively indicating the cessation of net structural growth.

> Figure 3: the two shades of orange used for the two scenarios are difficult to distinguish. A more contrasting color scheme would help.

The color scheme for Figure 4 (previously Figure 3) has been revised to improve distinguishability between scenarios.

> Section 3.3 would benefit from a graphical time series showing the onset of functional decline followed by structural changes (e.g., LAI). This would make the early warning argument much more tangible for the reader.

We thank the reviewer for this suggestion. Figure B5 has been added to the supplement showing standardized 15-year rolling mean trajectories of functional (NEP, WUE) and structural (LAI, AGB) variables for two representative SSP3-7.0 models, chosen to bracket the range of functional-to-structural lags across the ensemble (mri-esm2-0: 25 yr; gfdl-esm4: 12 yr). The figure reveals a three-tier early warning sequence in mri-esm2-0: NEP and WUE decline begins in 2037, LAI decline follows in 2044, and AGB stagnation is not detected until 2062, providing up to 25 years of functional warning before structural change becomes apparent in inventory-based metrics. In gfdl-esm4 all functional and structural growth variables decline simultaneously in 2037, with AGB stagnation following 12 years later, illustrating the model uncertainty in early warning window length. We refer to the figure in the results section.

L413–414: The temporal sequence of decline onset is illustrated for two representative models in Figure B5, showing functional decline detectable up to 25 years before AGB stagnation under SSP3-7.0.