Reviewer 1:

We would like to sincerely thank the reviewers, Rene Orth and Josephin Kroll, for their helpful and constructive comments. We found the feedback very valuable in improving and sharpening the manuscript and have revised it accordingly. Please find below our detailed point-by-point response in red, with the reviewer's comments for clarity. Line numbers refer to the revised manuscript version.

Reviewers #1: Rene Orth and Josephin Kroll.

Review of Heselschwerdt et al., egusphere-2025-5896

“Large impact of extreme precipitation on projected blue-green water shares“

This study investigates the partitioning of precipitation into runoff and transpiration, and its changes through this century. The authors introduce a new metric to describe this partitioning and use output of Earth system model simulations to determine the partitioning and its changes until the end of the century throughout the globe. They identify hot spot regions where precipitation is mainly partitioned into runoff or transpiration, respectively, and regions with strongest projected changes in the partitioning. Finally, they attribute the partitioning changes mainly to changes in extreme precipitation and vegetation and derive management implications.

Recommendation:

We think the paper requires major revisions.

The topic of this study is interesting and timely. Ongoing global change affects climate and vegetation in various ways, and their complex interplay modulates the water cycle. This way, major water fluxes such as runoff and transpiration may change differently, and for partly different reasons. The precipitation partitioning metric introduced by the authors in this paper summarizes relevant water cycle changes beyond individual variables, including an attribution of the resulting spatial patterns of change. This analysis presents a relevant contribution because it helps to identify regions with relevant water cycle changes, and proposes related mechanisms that require additional attention in further research.

However, before the paper can be published we would ask the authors to consider the concerns described below:

General comments:

The introduction section falls short in introducing the knowledge gap that this study addresses. This comes in combination with an insufficient review of existing literature on the topic. The present text claims that not much research exists in this direction (lines 62-63) while there is actually quite a body of literature that jointly analyses runoff and (evapo)transpiration, such as for example the Orth & Destouni 2018 study which is cited later, among others.

We have revised the Introduction to better acknowledge the existing literature on joint runoff and evapotranspiration/transpiration responses and on precipitation partitioning (Introduction, l. 46-56). In particular, we now cite Orth and Destouni (2018) already in the Introduction and frame our contribution more specifically as a global assessment of projected runoff-transpiration partitioning shifts, their dominant climatic and hydroecological controls, and their co-occurrence with absolute runoff and transpiration changes.

Related to this, lines 34-61 provide interesting mechanistic background on precipitation partitioning, which, however, is not connected to the rest of the introduction. So we suggest either to shift this to the discussion section, or connect it more to the introduction to motivate the relevance and knowledge gap in the topic.

We thank the reviewer for this suggestion. We have revised the introduction to more clearly connect the mechanistic background to the study aim (Introduction, l. 38-45). The discussion of plant physiological responses, vegetation change, atmospheric demand, and precipitation characteristics now directly motivates the driver analysis and clarifies why these controls are included in the attribution framework.

While the authors choose transpiration as green water flux, this choice is limiting the scope of this global analysis. For example, in arid areas transpiration is negligible while evaporation constitutes most of the surface water flux which is then ignored in the analysis. Additionally, soil moisture is referred to as green water and is modulated by transpiration but also evaporation.

We have revised the framing and methods accordingly. The analysis is now explicitly restricted to vegetated and hydrologically active land areas (Abstract, l. 2;

Introduction, l. 32-37; Sect. 2.4, l. 119-127). To make this scope restriction explicit and exclude very dry and sparsely vegetated regions, we now apply a fixed historical mask derived from the ensemble mean over 1985–2014, retaining only grid cells where mean precipitation exceeds 0.05 mm d^{-1}, mean runoff 0.005 mm d^{-1}, mean transpiration 0.005 mm d^{-1}, and mean LAI 0.3 $m^2 m^{-2}$ (Sect. 2.3, l. 104-110) . BGWS is now described more explicitly as a flow-based partitioning metric for runoff and transpiration rather than as a complete representation of all blue and green water components. We have also clarified why we focus on transpiration as the direct vegetation water-use response, while noting that evapotranspiration (ET) includes non-transpirational evaporation and therefore reflects a broader land-surface water-loss signal. In this sense, using ET instead of transpiration would address a different question, namely runoff versus total evaporative loss rather than runoff versus vegetation-mediated water use. In addition, we complement the main analysis by computing non-transpirational (nt) evaporation as $E_{nt} = ET - E_t$ and show its historical and projected changes in the Supplement (Eq. 3 in Sect. 2.4; Figs. S5c and S9c).

A large part of the explanation of the results and the respective discussion, as well as attribution is based on changes in vegetation characteristics. However, at least one model has prescribed LAI dynamics (CNRM-ESM2-1, Seferian et al 2019, see page 4186), which means that its vegetation dynamics may be underestimated and related to this also its contribution to blue-green water partitioning and related trends. We suggest a review of the current model selection in this context.

We have revised the model selection with a specific focus on LAI dynamics and vegetation representation (Sect. 2.1, l. 77-78; Table S1). The original 11-model set included CNRM-CM6-1, for which LAI is prescribed and land cover is fixed. We therefore removed CNRM-CM6-1 from the analysis. We retained CNRM-ESM2-1, because in this model LAI and phenology are interactive and follow the leaf carbon balance rather than a prescribed LAI climatology (Seferian et al., 2019).

We have also updated the ensemble from 11 to 12 models due to changes in the availability of model output (Table S1). We replaced CMCC-CM2-SR5 with the Earth system configuration CMCC-ESM2, and we added EC-Earth3-Veg and GFDL-ESM4 (Cherchi et al., 2019; Lovato et al., 2022; Döscher et al., 2022; Dunne et al., 2020). This gives a larger and more consistent ensemble for a study that discusses vegetation effects. CMCC-ESM2 extends terrestrial biogeochemical processes relative to the CMCC climate model configuration, EC-Earth3-Veg includes the dynamic vegetation model LPJ-GUESS, and GFDL-ESM4 represents vegetation dynamics and demography in GFDL-LM4.1. The revised 12-model ensemble is therefore better suited to the aims of this study than the original 11-model set.

We have also revised the model table (Table S1) to outline vegetation representation in the individual models. The retained models do not all simulate vegetation cover in the same way. A subset includes dynamic vegetation cover or demography, while others mainly represent vegetation through time-varying LAI and phenology.

Furthermore, while extreme precipitation is identified as a main control of trends in blue-green water shares, this may not be well represented by Earth system models. For example, a recent study found that the current rather coarse resolution of data from climate and earth system models, as also used in this study, underestimates the impact of such extremes (Brunner et al. 2025).

We agree and have now discussed this limitation more explicitly (Sect. 3.3, l. 348-351). We cite Brunner et al. (2025) in the revised discussion and note that coarse-resolution ESMs tend to underestimate precipitation extremes and often exhibit drizzle bias, so the hydrological influence of RX5day in our analysis is likely conservative.

While we appreciate the comprehensive set of potentially influential variables included in the regression analysis, we would like to suggest some additions. In addition to RX5day, where the chosen time scale is actually not motivated, you could test RX1day to capture even more extreme precipitation. Moreover, the seasonality of precipitation and VPD (even mentioned in line 407, could be expressed as e.g. standard deviation across monthly values) could influence the partitioning which is expected to be different in regions with pronounced rainy seasons versus regions where precipitation is more equally distributed across the year. Also land use change could play a role in some regions. While this is indirectly captured in leaf area index it could be more explicitly included by considering the time-varying crop and forest cover fractions. Further, assessing the role of near-surface soil moisture in addition to that of total-column soil moisture may be insightful because they can be expected to be of different relevance to runoff and transpiration. In this context, it is unclear which “relative changes” of soil moisture (line 99) are evaluated instead of actual soil moisture output, and why.

We thank the reviewer for these helpful suggestions. We have expanded the candidate predictor set and explicitly tested RX1day, precipitation seasonality, VPD seasonality, near-surface soil moisture, and land-use fractions (tree and crop cover, where available) (Sect. 2.5, l. 156-160; Sect. 2.6, l. 196-209; Supplement S2).

Based on these tests, we retained precipitation seasonality, VPD seasonality, and near-surface soil moisture in the final main predictor set. We also tested RX1day as an alternative extreme-precipitation metric and document this sensitivity analysis in Supplement S2 (S2, l. 793-797; Table S3). RX1day yielded only marginal improvements in mean out-of-sample performance relative to RX5day and did not substantially alter the dominant predictor ranking. We therefore retained RX5day in the main analysis because multi-day accumulation metrics show slightly better spatial agreement than single-day extremes in CMIP6 and observational evaluations, making RX5day the more robust large-scale indicator (Li et al., 2021; Dunn et al., 2022) (Sect. 2.6, l. 201-204). Single-day extremes are also generally more sensitive to coarse model resolution and therefore more uncertain in CMIP6-scale datasets (Brunner et al., 2025).

We further tested the contribution of land-use change by adding tree and crop cover fractions in a reduced 9-model ensemble for which these variables were available (Sect. S2, l. 798-802; Table S3). Adding these predictors led to only small performance changes relative to the main 12-model analysis, so we did not retain them in the main attribution and instead report them as sensitivity results. We have also clarified that LAI reflects both climatic and human influences and therefore partly captures land-use and land-cover change effects, although not explicitly (Sect. 2.6, l. 206-208). A more direct assessment requires a dedicated analysis; we now note this in the Conclusions (l. 452-454), and such an analysis is the focus of a subsequent study (Heselschwerdt et al., 2026).

We have also replaced total-column soil moisture in the main predictor set with near-surface soil moisture. The previous use of relative changes in total-column soil moisture was motivated by differing soil-column depths across ESMs, which limit the comparability of absolute values in a multi-model analysis.

The validation of the modelled partitioning against reanalysis and observation-based products in Figure S1 is informative. At the same time you are studying changes in the partitioning over time which is different from the partitioning itself; for this reason you could consider comparing changes in the partitioning during the observational period between models and reference products.

Further, Figure S1 is mentioned several times in section 3.1 such that it could be considered to move it to the main manuscript.

We thank the reviewer for this suggestion. In addition to the climatological benchmark shown in Supplement Fig. S1, we have now added a comparison of recent historical BGWS changes (2000–2014 minus 1985–1999) between the CMIP6 ensemble mean and the reference datasets to the Supplement (Fig. S2). This additional change-based benchmark is described in Sect. 2.2 (l. 96-98) and briefly discussed in Sect. 3.1 (l. 230-231). We note, however, that this short-period comparison is only a limited consistency check, since differences between two consecutive 15-year periods can be strongly affected by internal variability and are not directly comparable to the forced end-of-century response assessed from 30-year climatologies. We retained both benchmarking figures in the Supplement (Fig. S1 for historical mean BGWS and Fig. S2 for recent historical BGWS changes) to keep the main manuscript focused on the process interpretation and projected changes.

Related to this, we appreciate the indication of model uncertainty in Figures 2-4. However, this is not really visible in the maps in Figures 2b-d and 4, and missing in Figure 1. Maybe consider to make this more prominent - or even to introduce a masking to avoid showing results in regions where (i) models disagree or (ii) models disagree with reference products in Figure S1, in order to avoid interpreting results in regions where uncertainties actually largely prevent this. This could also enhance the accuracy of the attribution analysis in Figure 3.

We thank the reviewer for this suggestion. We have strengthened the uncertainty communication in several ways. Figure 1 now marks regions where the sign of the ensemble-mean BGWS disagrees with both reference datasets, Fig. 2 uses a stricter low-agreement threshold (<8/12 models agree on the sign of change), and Fig. 4 masks regions with low ensemble sign agreement (<8/12 models agree on the sign of change) entirely. In addition, we made the uncertainty scatters in Fig. 1 and 2 more visible.

For the attribution analysis, we additionally tested a version restricted to grid cells with robust inter-model sign agreement in ΔBGWS.

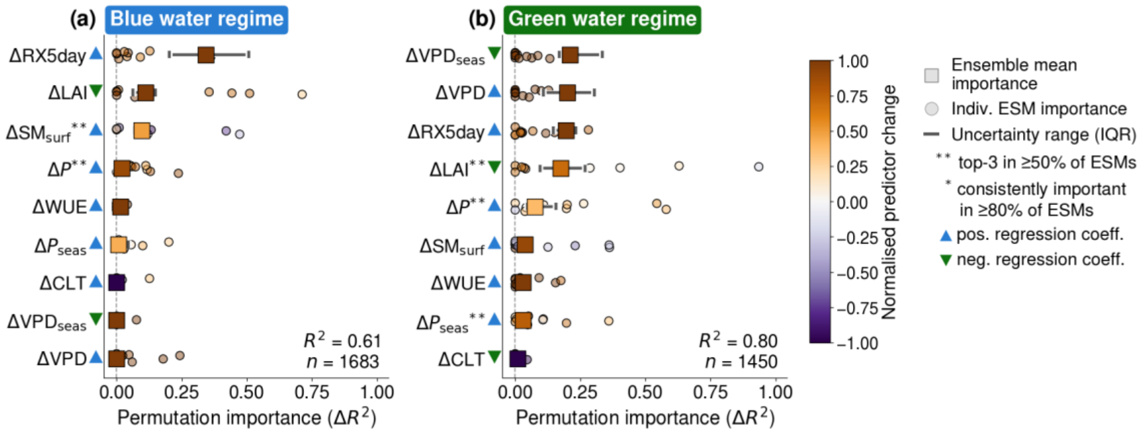

Figure 1. As Fig. 3 in the revised manuscript but restricted to grid cells where at least 8 out of the 12 ESMs agree on the sign of change.

Although this increased predictive skill, it did not substantially change the dominant predictors. Restricting the analysis in this way would also substantially reduce the number of grid cells and limit the attribution to only the more robust subset of each regime, which concentrates on climatic regions with lower model spread. We therefore retain this masked analysis as a sensitivity test rather than as the main analysis, so that the main attribution remains regime-wide and comparable across the full spatial domain.

We do not wish to remain anonymous - joint review by Rene Orth and Josephin Kroll.

We have notified the editor of our collaborations with Peter Greve and Lan Wang-Erlandsson.

Specific comments:

Line 4: given that this is the abstract, maybe replace blue/green water shares with a more generally understood terminology

We have revised the abstract to use more explicit partitioning terminology and now define blue and green water flows directly as runoff and transpiration (l. 1).

Line 4: Climate simulations → Earth system model simulations (this also applies to other mentions of climate simulations throughout the manuscript)

We have changed this wording throughout the manuscript as suggested.

Line 8: according to the results in Figure 3 it is leaf area index and water use efficiency individually rather than their interaction that influences the partitioning

We have revised the abstract to avoid implying an interaction term. In the revised analysis, water-use efficiency no longer ranks among the main predictors, whereas leaf area index is the second most important predictor across both regimes. We therefore now emphasise the role of leaf area index more clearly in the abstract (l. 8-9).

Line 12-16: this motivation of the relevance of precipitation partitioning is very brief, would expand this

We thank the reviewer for this comment. We have expanded the motivation in the opening paragraph to more clearly explain the relevance of precipitation partitioning for future water availability (introduction, l. 13-22).

Line 21: "stabilises Earth system" is unclear and

Line 23: "climate regulation" is unclear and

Line 27: "critical water functions" is unclear and

Line 54: "vegetation activity" is unclear and

Line 57: "runoff-consuming sinks" is unclear

We have revised the introduction to remove these unclear formulations. We now describe the relevant ecosystem, land-atmosphere, and water-cycle functions more explicitly, and we use more precise terms for vegetation and runoff-related processes.

Line 80: required variables are unclear at this point

We have clarified more clearly that the ESMs are selected so they provide the variables required to compute the BGWS metric and the climatic and hydroecological variables, and reference the relevant subsections here (Sect. 2.1, l. 76-77).

Line 114: Good point about the non-closure of the water balance. See our recent paper Huang et al. for this. (sorry for the self-promotion)

We thank the reviewer for this reference. We have now cited Huang et al. (2025) when discussing the non-closure of the water balance in the observation-based datasets (Sect. 2.2, l. 89-90).

Line 120: The term ‘climate indices’ does not fit the chosen variables/indices. For example, (i) RX5day is not calculated over a period that would refer to climate (~30 years), (ii) VPD is not an index, and also not referred to over a period considered as climate, (iii) WUE is not representing climate, but rather vegetation characteristics. An alternative could be ‘hydroecological variables’ as used in the manuscript before

We agree that the term “climate indices” was imprecise for the full variable set, because not all selected variables are indices and some represent hydrological or vegetation-related conditions rather than climate alone. We have therefore revised the terminology and now refer to them as “climatic and hydroecological variables” in Sect. 2.5. Here, “climatic” refers to the climatic domain, including precipitation, RX5day, VPD, and cloud cover, rather than implying that each variable is itself a 30-year climate index. RX5day is computed annually and then averaged over the historical and future 30-year periods. The hydroecological variables include soil moisture, LAI, and WUE. The subsection title and corresponding text have been updated accordingly.

Line 163: Explain more clearly in this section that this attribution analysis is done for the spatial patterns rather than for the temporal changes in each grid cell or region. Also emphasize this in the discussion of Figure 3 that you are not directly attributing the temporal changes of the partitioning shown in Figure 2 but merely the spatial patterns therein, where determined drivers are not necessarily the same for both.

We have revised the methods to state explicitly that the regression explains the spatial pattern of projected ΔBGWS across grid cells and does not attribute the temporal evolution at individual grid cells or regions (Sect. 2.6, l. 178-182). We also reiterate this framing in the discussion of Fig. 3 (e.g., Sect. 3.3, l. 337)

Line 170: Would start this section with this paragraph.

We have reordered the subsection so that it now begins with the general framing of the spatial attribution analysis (Sect. 3.3).

Line 192: “destroyed” → ”removed” and

Line 192/193: not sure we get the point here

We have clarified this point by simplifying the explanation of permutation importance and removing ambiguous wording (Sect. 3.3, l.210-216).

Line 195: Can you comment on the consideration of lateral flow in the models, and its potential role for the partitioning?

We added that lateral groundwater redistribution is often not represented or strongly simplified in many large-scale land models used in ESMs, in part because coarse model resolution does not resolve subgrid land-surface heterogeneity well (Liao et al., 2025). This may contribute to a blue-biased BGWS where shallow groundwater support to evapotranspiration is underrepresented. In the revised manuscript, we therefore added a short note in Sect. 3.1, where the historical BGWS bias relative to the reference datasets is discussed, stating that underrepresentation of shallow groundwater support to evapotranspiration may partly contribute to the underestimated E_t/P and thus to the blue bias (Maxwell and Condon, 2016) (Sect. 3.1, l. 234-237). We also added a limitation statement in the Conclusions accordingly (l. 446-450).

Line 199-201 & 326 & elsewhere: The results section incorporates interpretation/discussion already such that it would be more clear to adapt the title to 'Results and Discussion' and remove 'Discussion and' from the title of section 4.

We have adapted the section titles accordingly.

Line 202: unclear what's meant with absolute BGWS values and whether it is referring to model-mean data or observation-based results

We have changed it to "ensemble mean BGWS values" to make clear that this refers to the absolute BGWS values of the CMIP6 ensemble mean in comparison with the reference products (Sect. 3.1, l. 231-232).

Line 207: why interpret regions with large baseline biases at all?

We now avoid detailed interpretation in regions where the ensemble mean disagrees with both reference datasets and indicate these regions explicitly in Fig. 1.

Line 209: Would start the section with this paragraph.

We have reorganised this section so that it now opens with a higher-level summary of the main historical BGWS features (Sect. 3.1, l. 219-223).

Line 215: Unclear to which average 'below-average' is referring - global?

We have changed the sentence and clarify that it refers now to below-global-average E_t/P (Sect. 3.1, l. 245).

Lines 231/232: There are no visible patterns in the Sahara. and

Line 233: unclear what's the difference between saturation excess vs. infiltration excess

We have revised this passage by removing the Sahara example and instead discussing retained vegetated drylands more generally (Sect. 3.1, l. 258-266). In the revised analysis, hyper-arid and sparsely vegetated regions are excluded by the fixed historical mask, so the Sahara is no longer part of the analysed domain. We have also removed the previous sentence on saturation- and infiltration-excess runoff from this location.

Line 244: Can/Should you assume broadly similar soil properties across rainforests?

We removed the statement assuming broadly similar soil properties across rainforests. The revised text no longer relies on this assumption in explaining the contrasting historical BGWS patterns across tropical rainforest regions (Sect. 3.1, l. 269-274).

Line 257: section 3.2 seems a bit convoluted and was harder to follow: suggestion to rather summarize the main possible mechanisms and then mention example regions instead of selecting regions and describing those

We have restructured Sect. 3.2 to emphasise the main process pathways first and then use regional examples to illustrate them.

Line 265: Fig. 2: place legend in panel a on top or below as for the other figures; brings space for spatial average subpanel

We have revised Figure 2a and added an inset summarising the land-area fractions of the four BGWS-change classes.

Line 268: 'strongly linked' suggests causality, which cannot be derived from covariance across space

We have revised this wording throughout to avoid implying causality from spatial covariance and now use more cautious formulations.

Line 342: increases instead of ‘is increasing’

We have corrected this wording.

Line 351: blue-to-green shift is of similar magnitude, why is this not covered in the text?

We now discuss this more explicitly in Sect. 3.3 by highlighting that larger LAI increases provide the main counterweight to the blueward influence of stronger precipitation extremes and help explain why blue-to-green shifts still occupy a substantial fraction of global land area (Sect. 3.3, l. 371-373).

Line 352: Actually this section is about implications rather than management. There is little discussion about land or water management measures and strategies. Either revise this section, or drop the management from the title here, and also adapt the wording in the abstract in line 9 (“impact-relevant”, “actionable”).

We agree and have revised this section accordingly. The section is now titled 'Implications of future BGWS trends', and we have reduced the management framing in both the main text and the abstract.

Line 355: One could ask why it matters where the next rain drop goes in the presence of little to no rain in some regions

We now address this directly in Sect. 3.4 by stating that the analysis excludes hyper-arid and sparsely vegetated regions and by framing BGWS as most relevant for impacts related to hydrological sensitivity and timing (Sect. 3.4, l. 396-397).

Line 368: insert Δ before changes as done in the lines below

We have corrected this notation.

Line 385: I wonder about the causal direction here: are transpiration increases supporting vegetation productivity or does increased vegetation productivity induce/come with increased transpiration?

We revised the wording to avoid implying a single causal direction between transpiration and vegetation productivity. The revised text now describes these cases more cautiously in terms of greener partitioning and stronger latent cooling (Sect. 3.4, l. 418-420).

Line 387: Figure 5: this is rather a table than it is a figure. For a figure, I would suggest to make two x-y graphics, having dR or dT on the y-axis and dBGWS on the x-axis. Insert text in the respective 4 panels of the graph

We have substantially revised this figure. The previous Figure 5 has been replaced by a new Figure 4 that combines the robust sign combinations of ΔBGWS with ΔR and ΔE_t in map form and includes a schematic summary table of the corresponding interpretations.

Line 392: 'sizeable areas' is unclear and

Line 398: 'environmental flows' is unclear

We have removed this wording and revised the passage to use more precise language. The text now states more clearly that regionalised and observationally constrained analyses would still be needed before drawing site-specific management conclusions (Sect. 3.4, l. 427-430).

- Figure 1: The content of section 3.1 mainly describes and discusses results in the light of the i) water-energy-limitation framework as well as ii) the role of extreme precipitation. Hence, Figure 1 should include maps illustrating the three BGWS features - energy-limited partitioning, water-limited partitioning, precipitation-intensity related partitioning - and the ratio of RX5day/mean precipitation (so fig. S5a). This would make the figure more in line with the title of the manuscript, which emphasizes the role of extreme precipitation and follows the explanation used in the text. Finally, it would be good to cite some more related literature such as Seneviratne et al., 2010 and Denissen et al. 2022.

We have revised Figure 1 to better reflect the hydroclimatic interpretation of the historical BGWS patterns. The revised figure now includes an inset hydroclimatic regime classification, in which each grid cell is assigned to the dominant standardized signal of low near-surface soil moisture, low air temperature, or high precipitation seasonality. The text now frames the three large-scale features as energy-limited, water-limited, and precipitation-seasonality regulated partitioning. We have also added the suggested references to Seneviratne et al. (2010) and Denissen et al. (2022) (Sect. 3.1, l. 220-223). We retained the RX5day ratio maps as the Supplement Fig. S6 rather than moving them into the main figure.

- Figures 1 and 2: Adjust color bar range across panels b-d in both figures.

We have adjusted the colour scales in the revised figures to improve cross-panel comparability. In Figure 1, R/P and E_t/P now use matching percentage scales, and in Figure 2, ΔP, ΔR, and ΔE_t use a common range.

- Figure 3: Do the boxes in panels a and b show the spread of the attribution results for the individual Earth system models? Also, panels c and d are not described and discussed in the text. For example, (i) y-axis unclear, response to what?, and (ii) what is the meaning of the lines connecting 'end-of-century response' across predictors, especially as the y-axis scale changes. Instead, you could use boxplots and depict individual models by differently colored dots or differently shaped points; +/- dBGWS could be indicated by outline of each dot/symbol or by making two boxplots per predictor.

We have completely revised Figure 3. The previous panels c and d, which used a parallel-coordinates plot for displaying multivariate model responses (e.g. Mankin et al., 2019), were removed. The revised figure now focuses on the permutation importance of the ensemble mean, shows the spread across repeated blocked cross-validation splits, and additionally overlays permutation importance from the individual ESMs.

- Figure S1: Labelling of panels in the caption is not in line with the figure.

We have corrected the caption labelling.

- You could consider adding summary statistics to the maps in the figures as also stated in the text.

We have added summary statistics to the revised figures where most useful, including inset area fractions in Fig. 2, land area fractions in Fig. 4, agreement fractions in Fig. S1 and S2, and land area fractions in Fig. S8.

References:

Orth, R. and G. Destouni 2018: Drought reduces blue-water fluxes more strongly than green-water fluxes in Europe. https://doi.org/10.1038/s41467-018-06013-7

Huang, H. et al. State-of-the-art hydrological datasets exhibit low water balance consistency globally, doi: 10.5194/essd-2025-376

Séférian et al. 2019: Evaluation of CNRM Earth System Model, CNRM-ESM2-1: Role of Earth System Processes in Present-Day and Future Climate. https://doi.org/10.1029/2019MS001791

Seneviratne et al. 2010: Investigating soil moisture-climate interactions in a changing climate: A review. https://doi.org/10.1016/j.earscirev.2010.02.004

Denissen et al. 2022: Widespread shift from ecosystem energy to water limitation with climate change. https://www.nature.com/articles/s41558-022-01403-8

Brunner et al. 2025: A global perspective on the spatial representation of climate extremes from km-scale models. https://doi.org/10.1088/1748-9326/ade1ef

New References:

Cherchi, A., Fogli, P. G., Lovato, T., Peano, D., Iovino, D., Gualdi, S., Masina, S., Scoccimarro, E., Materia, S., Bellucci, A., and Navarra, A.: Global mean climate and main patterns of variability in the CMCC-CM2 coupled model, J. Adv. Model. Earth Syst., 11, 185-209, https://doi.org/10.1029/2018MS001369, 2019.

Döscher, R., Acosta, M., Alessandri, A., et al.: The EC-Earth3 Earth system model for the Coupled Model Intercomparison Project 6, Geosci. Model Dev., 15, 2973-3020, https://doi.org/10.5194/gmd-15-2973-2022, 2022.

Dunne, J. P., Horowitz, L. W., Adcroft, A. J., et al.: The GFDL Earth System Model version 4.1 (GFDL-ESM 4.1): overall coupled model description and simulation characteristics, J. Adv. Model. Earth Syst., 12, e2019MS002015, https://doi.org/10.1029/2019MS002015, 2020.

Dunn, R. J. H., Donat, M. G., and Alexander, L. V.: Comparing extremes indices in recent observational and reanalysis products, Frontiers in Climate, 4, https://doi.org/10.3389/fclim.2022.989505, 2022.

Heselschwerdt, S. P., Dengri, A., Fahrenbach, N. L. S., and Greve, P.: Impacts of land-use and land-cover change on blue–green water partitioning, EGUsphere [preprint], https://doi.org/10.5194/egusphere-2026-513, 2026.

Li, C., Zwiers, F., Zhang, X., Li, G., Sun, Y., and Wehner, M.: Changes in Annual Extremes of Daily Temperature and Precipitation in CMIP6 Models, Journal of Climate, 34, 3441-3460, https://doi.org/10.1175/JCLI-D-19-1013.1, 2021.

Liao, C., Leung, L. R., Fang, Y., Tesfa, T., and Negron-Juarez, R.: Representing lateral groundwater flow from land to river in Earth system models, Geoscientific Model Development, 18, 4601–4624, https://doi.org/10.5194/gmd-18-4601-2025, 2025.

Lovato, T., Peano, D., Butenschön, M., Materia, S., Iovino, D., Scoccimarro, E., Fogli, P. G., Cherchi, A., Bellucci, A., Gualdi, S., Masina, S., and Navarra, A.: CMIP6 simulations with the CMCC Earth System Model (CMCC-ESM2), J. Adv. Model. Earth Syst., 14, e2021MS002814, https://doi.org/10.1029/2021MS002814, 2022.

Maxwell, R. M. and Condon, L. E.: Connections between groundwater flow and transpiration partitioning, Science, 353, 377–380, https://doi.org/10.1126/science.aaf7891, 2016.

Reviewer 2:

We thank the reviewer for the positive overall assessment of our study and for the constructive comments and suggestions. We found the feedback very valuable and have revised the manuscript accordingly. Below, we provide a detailed point-by-point response in red and indicate where the manuscript has been updated. Line numbers refer to the revised manuscript version.

The authors use the results of 11 Earth System Models to estimate the current partitioning of precipitation into runoff and transpiration and how this partitioning changes under the SSP3-RCP7.0 combined socioeconomic scenario. They also determine which climatic and ecophysiological factors could be responsible for any changes using multiple linear regression methods with analyses of predictor importance (size of regression coefficient and randomnization of predictors). This is an interesting study that deserves publication. I do have some major and some minor issues:

Major issues

1. **The use of green and blue water. The BGWS metric is more a measure of process than a measure of quantity fir for management decisions (see hereafter). As such I do not see why one should use the terms green and blue water that are from sustainability water science, rather than from hydrology, which is about process. It would have been much easier to look at a hydrological measure such as the runoff coefficient RC and its change. 1 -RC is then the change in evaporation which is directly related to transpiration as well.**

We agree that BGWS is primarily a process-based indicator rather than a quantity-based metric for direct management decisions. We have clarified this more explicitly in the revised manuscript, reduced the management language, and reframed the respective section towards implications rather than management (Sect. 3.4).

We have also clarified why we retain blue-green water language (Introduction, l. 23-31). While our analysis is fundamentally about runoff-transpiration partitioning, we use the blue-green water framework as an impact-oriented framing because it links hydrological partitioning to its different relevance for ecosystems and human water use, rather than describing the hydrological process alone. The revised Introduction now distinguishes blue water stores and flows from green water stores and flows, and defines our operational use of blue-green water partitioning as the partitioning of precipitation into runoff and transpiration.

Finally, we have clarified why we do not use the runoff coefficient, or its complement, as the main metric (Introduction, l. 32–37). We agree that runoff-coefficient-based metrics provide a useful hydrological perspective on runoff or the broader non-runoff share of precipitation. However, they do not isolate vegetation-mediated water use. Our study is specifically interested in the balance between runoff and transpiration, because transpiration captures a key pathway of plant water use. We now state this explicitly in the Introduction, noting that metrics based on ET or the runoff coefficient describe broader water-loss or non-runoff signals and do not isolate vegetation-mediated water use. We also clarify in Sect. 2.4 that BGWS is a flow-based partitioning metric for runoff and transpiration, not a complete representation of all blue and green water or water-balance components, and that it is complementary to existing partitioning metrics.

2. **If water management is the main target, why not look at absolute changes in runoff (Q) and transpiration (T). If I live in a water-stressed region, I would be happy if both Q and T increase, regardless of whether or the other changes more to lead to a change in the BGWS. In section 3.4 the authors themselves show that for interpretation for water management, one requires to take along changes in Q (floods) or T (drought). This seems overly cumbersome if we can look at changes in P, Q and T directly.**

We agree that absolute changes in runoff and transpiration are more directly relevant for water management than BGWS alone. Water management is not the main target of this study. In the revised manuscript, we therefore reduced the management language and reframed Sect. 3.4 towards implications rather than management. We now make clearer that BGWS is a process-based indicator that is most informative when interpreted jointly with absolute changes in runoff and transpiration.

This distinction is now stated explicitly in Sect. 3.4. There, we explain that BGWS addresses the question of where the next unit of rain tends to go, rather than how much water is available overall. In this sense, BGWS complements absolute flux changes by providing additional insight into hydrological sensitivity, partitioning behaviour, and the controls of future change. This also supports a process-oriented evaluation of ESMs that goes beyond comparing absolute variables alone (Sect. 3.4, l. 393-401).

We therefore do not view BGWS as a substitute for analysing absolute changes in precipitation, runoff, or transpiration. Rather, we use BGWS alongside these fluxes to link absolute changes with changes in runoff-transpiration partitioning and their broader hydroecological implications, as shown in Fig. 4.

3. **I wonder also how much of the changes in BGWS should be attributed to land cover change. If the ESMs are driven by combined SSP-RCP scenarios, exogeneous land cover change is likely included in the ESM land modules, which has a direct impact on evaporation and runoff (see e.g. Bosmans et. al, 2017: https://hess.copernicus.org/articles/21/5603/2017/). Also, human water use change such as the extension of in irrigated areas, when included in the land modules, will play a role.**

We agree that land-cover and other human land-surface changes can influence blue-green water partitioning. We therefore explicitly tested tree and crop cover fractions in a reduced sensitivity analysis for the subset of models where these variables were available (Supplement S2, l. 798-802; Table S3). Including these predictors led to only small performance changes, while reducing the ensemble size from 12 to 9 models. We therefore did not retain them in the main attribution and instead report them as sensitivity results.

We have also clarified in the revised manuscript that LAI likely reflects both climatic and human influences and therefore partly captures land-use and land-cover change effects, although not explicitly (Sect. 2.6, l. 206-208). In the Conclusions, we now state that future work should assess land-use and land-cover changes more directly (l. 452-454). A more dedicated analysis of these effects is the focus of a subsequent study (Heselschwerdt et al., 2026). We also note there that improving the treatment of human influences such as irrigation remains important for more reliable future projections, as these processes are still simplified or not at all represented across CMIP6 ESMs (Conclusions, l. 454-456). We are convinced that a deeper assessment of land-use and land-cover change impacts would require a more dedicated analysis than is feasible within the scope of the present study.

4. **What is missing from the possible important factors is the impact of climate seasonality. For regions that have constant rainfall throughout the year and a clear summer season with more radiation (growing season determined by energy limitation), an equally proportional increase in temperature and precipitation will likely leadto a higher runoff coefficient, because almost all the increased precipitation will runoff in the winter and the increased temperature will not lead to a proportional increase in evaporation because of warm season water limitation. This is even more so the case in areas where precipitation falls predominantly in winter.**

We have addressed this by adding precipitation seasonality and VPD seasonality to the candidate predictor set and retaining both in the revised analysis (Sect. 2.5-2.6). In addition, Sect. 3.1 now frames one of the main historical BGWS features as precipitation-seasonality regulated partitioning, and Figure 1 includes a hydroclimatic regime inset based partly on precipitation seasonality.

Minor issues

1. **Global climate models and likely earth system models are often suffering from permanent drizzle, underestimating precipitation intensity. This could lead to an underestimation of the runoff coefficient. Also, the hydrology does not always contain a sub-grid parameterization that results in the correct saturation excess runoff, making runoff overly flashy. If this is a potential issue, this should be discussed.**

We have addressed this in Sect. 3.3. The revised manuscript now notes that coarse-resolution ESMs tend to underestimate precipitation extremes and often exhibit drizzle bias, which likely makes the hydrological influence of RX5day conservative in our analysis (Sect. 3.3, l. 349-351). We also clarify that the runoff responses are based on the simplified runoff-generation schemes used in CMIP6 land models, which mainly represent saturation-excess runoff and may underrepresent infiltration-excess processes (Sect. 3.3, l. 347-349).

2. **On the other side of the spectrum: it could be that the limited representation of groundwater convergence in ESMs underestimates the fact that shallow groundwater will support evaporation in drylands or during dry periods. This may partly explain the over-estimation of the BGWS as seen in Figure S1. See e.g. Maxwell and Condon (2016). https://www.science.org/doi/abs/10.1126/science.aaf7891**

We agree that lateral groundwater redistribution is often not represented or is strongly simplified in many large-scale land models used in ESMs, in part because coarse model resolution does not resolve subgrid land-surface heterogeneity well (Liao et al., 2025). This may contribute to a blue-biased BGWS where shallow groundwater support to evapotranspiration is underrepresented. In the revised manuscript, we therefore added a short note in Sect. 3.1, where the historical BGWS bias relative to the reference datasets is discussed, stating that underrepresentation of shallow groundwater support to evapotranspiration may partly contribute to the underestimated Et/P and thus to the blue bias (Maxwell and Condon, 2016; Sect. 3.1, l. 234-237). We also added a corresponding limitation statement in the Conclusions (l. 446-450).

Line 109 and line 113: GLEAM includes a soil moisture accounting model to reduce potential evaporation. As such it is a modelled product itself. It is better to compare evaporation of the ESMs with in-situ laten heat fluxes from e.g. FLUXNET (https://fluxnet.org/).

We agree and have revised the methods accordingly. We now state more explicitly that GPCC, G-RUN, and GLEAM are derived using different methodologies, that GLEAM is a hybrid observation-model product, and that these reference datasets are therefore not equally direct observations and are not fully consistent with one another or with a closed global water balance (Sect. 2.2, l. 86-90). We also clarify that this study uses global-scale gridded reference datasets for a first-order consistency check of how well the simulated BGWS agrees with selected reference products, rather than for a full benchmarking of ESM skill (Sect. 2.2, l. 94-96). A FLUXNET-based evaluation would require a different site-scale evaluation strategy and is therefore beyond the scope of the present study.

3. **Lines 212-213: Here, part of the BGWS is explained by energy limitation and Budyko. Then why not add a map of the aridity index to corroborate this and also use this as a predictor?**

We agree that the historical interpretation is conceptually consistent with Budyko and aridity-based thinking. To better corroborate this interpretation, the revised Fig. 1 now includes a hydroclimatic regime inset based on low near-surface soil moisture, low air temperature, and high precipitation seasonality, and Sect. 3.1 links the energy-limited BGWS pattern to low net radiation, low air temperature, below-global-average Et/P, and comparatively high near-surface soil moisture. We therefore use these direct hydroclimatic variables to support the interpretation rather than adding a separate aridity-index map.

We did not add aridity index as a separate predictor in the attribution analysis because it is a composite hydroclimatic diagnostic derived from underlying water- and energy-supply terms and would therefore be partly redundant with predictors already included in the model, particularly precipitation and atmospheric-demand related variables. Our intention is to analyse the contributing hydroclimatic controls directly rather than to add a second-order diagnostic that combines them. Similar reasoning has been used in other hydrological predictor-selection settings, where aridity index was excluded because it is deterministically derived from underlying climatic variables and may introduce redundancy in the analysis (Abbasizadeh et al., 2025).

4. **Lines 219-220: In high altitudes it is also likely that runoff is overestimated due to the underestimation of snow sublimation in ESM: see e.g.Stigter et al. (2018). (https://www.frontiersin.org/journals/earth-science/articles/10.3389/feart.2018.00108/full**

We agree and have added a short sentence in the discussion of mountain and cold-region BGWS patterns (Sect. 3.1, l. 247-251).

5. **Lines 248-253: Here the differences in BGWS between Southern China, Japan and Florida are discussed. Is topography a factor to consider here?**

We agree that topography may contribute to some of these regional contrasts. In the revised manuscript, however, this section now focuses more explicitly on the large-scale role of precipitation seasonality and uses the regional examples only illustratively (Sect. 3.1, l. 267–276). Although topography is not considered here explicitly, we note in Sect. 3.1 (l. 247–249) that orographically enhanced precipitation and topography-controlled runoff parameterisations can contribute to larger blue water shares in mountain regions.

References:

Abbasizadeh, H., Maca, P., Hanel, M., Troldborg, M., and AghaKouchak, A.: Can causal discovery lead to a more robust prediction model for runoff signatures?, Hydrol. Earth Syst. Sci., 29, 4761-4790, https://doi.org/10.5194/hess-29-4761-2025, 2025.

Brunner, L., Poschlod, B., Dutra, E., Fischer, E. M., Martius, O., and Sillmann, J.: A global perspective on the spatial representation of climate extremes from km-scale models, Environmental Research Letters, 20, 074 054, https://doi.org/10.1088/1748-9326/ade1ef, 2025.

Heselschwerdt, S. P., Dengri, A., Fahrenbach, N. L. S., and Greve, P.: Impacts of land-use and land-cover change on blue-green water partitioning, EGUsphere [preprint], https://doi.org/10.5194/egusphere-2026-513, 2026.

Liao, C., Leung, L. R., Fang, Y., Tesfa, T., and Negron-Juarez, R.: Representing lateral groundwater flow from land to river in Earth system models, Geoscientific Model Development, 18, 4601–4624, https://doi.org/10.5194/gmd-18-4601-2025, 2025.

Maxwell, R. M. and Condon, L. E.: Connections between groundwater flow and transpiration partitioning, Science, 353, 377-380, https://doi.org/10.1126/science.aaf7891, 2016.