RC2

This study examines the role of land use and land cover change on the partitioning of precipitation to blue (runoff) and green (transpiration) water fluxes. The authors determine how a switch between two different land use and land cover change (LULCC) scenarios (SSP1-2.6 and SSP3-7.0) alters blue-green water partitioning across four Earth system models from the Land Use Model Intercomparison project. The authors relate changes in blue-green water partitioning to vegetation functioning, considering varying sensitivity among the data distribution. They strongly emphasize and discuss inter-model spread throughout the manuscript.

Recommendation: The paper needs major revisions.

The paper is an important contribution to advance the understanding of how water fluxes will change in the future in response to land use and land cover change. The authors substantially widen the area across which the role of LULCC for water partitioning has been examined by using a so far only regionally applied methodology globally. The authors utilize data from CMIP6/LUMIP for their analysis and find substantial inter-model spread regarding the role of LULCC for future water cycle projections. This makes it hard to draw robust conclusions regarding the role of LULCC for blue-green water partitioning. However, it also enables the authors to provide an extensive discussion of possible reasons for model uncertainty and disagreement thereby granting insights for model development.

We thank the reviewer for the insightful and positive assessment. Please find our point-by-point responses below in blue, describing the planned revisions.

Before publication of the manuscript, the following points need to be considered:

Major comments:

1. We feel that the main focus of the paper on future changes in water and carbon cycling in response to land use and land cover changes (Figures 1-3) requires revision given the substantial differences between individual models in terms of the translation of land use/land cover changes into changes in LAI, GPP and BGWS. That said, a possible way forward could be to identify regions where the diagnosed signals in the four chosen models are relatively robust and representative, e.g. by determining regions where the four-model-ensemble signal agrees with that of the nine-model-ensemble in the historical period in terms of the sign (Figure S4). In other regions, however, results should not be shown (e.g. masking with grey color), and the discussion should focus rather on uncertainties (which it already does in many parts of the manuscript) than on diagnosed signals that can inform climate change adaptation and mitigation.

   We fully agree that robustness should govern what we show and claim, and we will adapt both the analysis and its interpretation accordingly. We will make the four-model sign-agreement ($\geq$3 of 4 models) visually dominant in Figs. 1–3, confine quantitative, adaptation-relevant statements to higher-agreement regions, and shift the interpretation further towards explaining where and why the response is uncertain, reframing the research questions accordingly (see comment 2). As this characterisation of inter-model disagreement is itself central to the study, we propose to mark the low-agreement regions clearly and discuss their uncertainty rather than removing them. We will additionally use the historical four- and nine-model comparison (Fig. S4) as a further robustness benchmark and discuss explicitly where it supports or limits confidence in the future signal. We use it as a benchmark, noting that the historical agreement need not carry over to the 2070–2099 contrast and that the four core models are part of the nine-model historical ensemble, so this comparison is not fully independent.

2. Related to the above comment, (i) the title currently does not reflect the main finding of the study where model uncertainty does not actually allow for robust conclusions of the impacts of land use and land cover change, (ii) the first two

research questions stated at the end of the introduction (lines 79-80) can not be answered given the uncertainty across models; and in fact these research questions are not revisited later in the manuscript because the authors acknowledge the uncertainty across models, thereby comprehensively addressing their third research question.

(i) We will revise the title to foreground the model-dependence of the response. (ii) We will rephrase the research questions so they are explicitly framed as a search for robust patterns across the available LUMIP ESMs rather than as questions with a single robust answer, and we will make clear that they are revisited and largely reframed around uncertainty in the Discussion. We agree that, as currently phrased, RQ1–RQ2 over-promise given the inter-model spread.

3. The authors discuss in the text related to Figures 1-3 that/which land use/land cover changes translate to changes in LAI, GPP and BGWS. At the same time, the figures only show resulting changes in LAI, GPP and BGWS while omitting the information on underlying land use/land cover changes. More generally, the motivation for studying LAI and GPP changes in Figures 1 and 2 is not clear. It would be informative to also show land use/land cover changes in the main figures in order to relate the resulting LAI, GPP and BGWS changes to this cause. We acknowledge that some land use/land cover change information is provided in the supplementary material, but this should be related to the content of main figures 1-3 and moved to the main material.

   We will move the LUH2 sustainable-LULCC forcing (the SSP1−SSP3 difference fields for forest, cropland, grazing land, and natural non-forest; currently Fig. S1i–l) into the main text, ahead of the LAI/GPP and BGWS figures, so the chain forcing $\rightarrow$ vegetation response $\rightarrow$ BGWS is visible there. We use the LUH2 forcing rather than the modelled land-cover diagnostics (Fig. S2) because it is complete and identical across models, whereas the latter are available only for some models and variables and use inconsistent definitions. The modelled response (Fig. S2) will be retained in the Supplement and be cited where we discuss the LUH2-to-model translation (uncertainty source 1) and in the eastern-US case study (Sect. 3.3). We will also explain why we analyse LAI and GPP rather than land-cover fractions: the same forcing maps onto different PFTs and tree-cover definitions across models, so cover fractions are not comparable, whereas LAI and GPP capture the functional vegetation response that drives transpiration.

4. Some methodological choices made by the authors required more justification. For example, the interpretation of the BGWS results is complicated by the fact that transpiration, which is chosen as a green water flux here, may be small compared to evaporation in some (arid) regions (line 167). Also, the thresholds applied for the amount of the investigated water fluxes are very small (line 177) such that weak signals may be overinterpreted such as changes in water flux partitioning in arid regions with little precipitation.

   We will adapt the activity thresholds, retaining only grid cells with mean precipitation above $0.822\,\mathrm{mm\,day^{-1}}$ ($\approx$300 $\mathrm{mm\,yr^{-1}}$) and mean runoff and transpiration each above $0.05\,\mathrm{mm\,day^{-1}}$, and report the resulting masked area. The precipitation threshold follows the order of magnitude of the IPCC definition of arid zones (IPCC, 2022) and the runoff threshold follows Schleussner et al. (2016). The same value is applied to transpiration as both fluxes enter BGWS symmetrically. This masks the arid regions where transpiration is usually small relative to evaporation and where the small precipitation denominator in Eq. (3) renders ΔBGWS unstable.

   We will also strengthen the justification for using transpiration rather than total ET as the green-water flux. Transpiration is the productive flux through which plants couple water loss to carbon uptake and water-use efficiency, with implications for vegetation productivity, evaporative cooling, and land–atmosphere exchange, whereas total ET also includes soil evaporation and interception that can vary independently of vegetation functioning. We will make this explicit in the Introduction and Methods and add a supplementary figure indicating where the choice of green water flux matters most.

5. The manuscript provides comprehensive discussion around possible reasons for model disagreement. This is a strength of the current manuscript. However, the multifaceted reasons for model disagreement make the related discussion complex and hard to follow. Therefore, we would suggest introducing a schematic figure illustrating the sources of model uncertainties and disagreement. Then, the text related to Figure 4 and the first part of the discussion section could be linked with this figure.

   We will add a schematic figure summarising the main sources of inter-model disagreement (LUH2-to-land-cover translation; PFT mapping and interactive competition; vegetation physiology and biogeochemistry, including nutrient limitation; and land energy–water and runoff parametrisation) and link it to the relevant text sections.

6. There is inconsistent terminology across the manuscript. For example, (i) change and anomaly are used interchangeably (e.g. in section 3.2), (ii) the difference between ΔLULCC and ΔsustainLULCC is not clear, (iii) IBGWS is not referred to in the results section, and (iv) vegetation water use is not defined while probably used as a description of transpiration.

   We will harmonise terminology throughout: (i) we will avoid using "change" and "anomaly" interchangeably. Because our signal is the difference between two land-use scenarios (L1 vs. L3) rather than a temporal change or a deviation from a climatological mean, we will consistently call it a "difference" for the quantity itself (e.g. ΔBGWS, ΔLAI) and a "response" when describing the reaction of the system to the LULCC forcing; (ii) since Δ_{LULCC} and $\Delta\mathrm{BGWS}_{\mathrm{sustain\,LULCC}}$ denote the same operator (generic vs. applied to BGWS), we will use a single, consistently defined notation throughout; (iii) we will refer to I_{BGWS} explicitly in the Results where its value (≈5.2 %) is reported; and (iv) we will define and retain "vegetation water use" only where a broader meaning is intended and use transpiration otherwise.

Minor comments:

– While the introduction section provides a useful background on the state-of-the-art of impacts of land use/land cover changes on water cycling, it does not describe the existing knowledge gap(s) which are addressed in this study. This makes it difficult to link this section with the actual analysis described in the further sections.

  We will add an explicit statement of the knowledge gap at the end of the Introduction — the scarcity of global, multi-model assessments that quantify the response of blue–green partitioning to LULCC and characterise inter-model spread — linking the background directly to our analysis.

– Research question 2 develops around the role of vegetation water use (= transpiration?) or precipitation feedbacks for changes in BGWS. Currently, the findings for this are hidden in the supplement and not combined. This could be nicely illustrated by a main figure combining categories from the change in ET or P with the change in BGWS.

  We agree this should be a main result. We will add a main-text categorical map classifying each grid cell by whether its ΔBGWS response is primarily associated with changes in transpiration, precipitation, or both, so RQ2 is answered explicitly in the main text rather than via the Supplement.

– We would suggest to combine the discussion with the results section rather than with the conclusion.

Given the length and the number of regional examples, we prefer to retain a separate Discussion but will improve signposting between Results and Discussion and add the subheadings requested by Reviewer 1.

- Line 8: the term "green water gains" is not defined

  We will define "green water gains" at first use (an increase in the green water share, i.e. a shift of partitioning toward transpiration / more negative ΔBGWS).

- Lines 26 & 56: the paper of Hoek van Dijke et al. 2022 which is mentioned in other places in the manuscript, could be referenced here.

  We will add the Hoek van Dijke et al. (2022) citation at both locations.

- Usage of T instead of ET, reasoning given in Line 35ff. not sufficient

  We will strengthen this justification; please see our response to Major comment 4 above, where we explain why transpiration (the productive, vegetation-controlled flux) is used and what the consequences of this choice are.

- Line 36: Orth & Destouni 2018 do not perform a model-based analysis but focus on observation-based datasets

  We will move this citation out of the "global model analyses" clause and cite it appropriately as observation-/dataset-based evidence.

- Lines 36-38: Either explain the BGWS metric more, or omit this here and introduce it only in the methods section

  We will remove the BGWS definition from the Introduction and introduce the metric only in the Methods (Sect. 2.3).

- Line 58: depress $\longrightarrow$ reduce

  We will replace "depress" with "reduce".

- Line 100: It would be good to check the contribution of both terms to make sure that not one is dominant about the other which would indicate limited representativeness of the findings with respect to background climate

  We will add a supplementary figure decomposing Δ_{LULCC} into its two contributions (the land-use contrast under the S1 and under the S3 background) to show that neither background dominates the combined signal.

- Line 113 Specify ‘For context’ ; maybe add on how land-use in this time period looked - e.g. in Europe land use /deforestation had its highest rates during 19th century

We will clarify what the historical comparison adds and briefly describe the historical land-use trajectory (e.g. that major deforestation in some regions predates 1985, so historical − hist-noLu over 1985–2014 captures comparatively recent transitions). Supporting land-use information will be added to the Supplement.

– Line 124: content from S1 might be more relevant for following the discussion than the technical details provided in T1

We will move the most decision-relevant process information from Supplement S1 (e.g. runoff-scheme type; prognostic vs. prescribed vegetation) into Table 1, so readers can follow the model contrasts without consulting the Supplement.

– Line 127: Explain the usage of surface soil moisture, instead of e.g. total column soil moisture. How representative is a 30-year mean considering the high variability in the upper 10cm of the soil?

We use *mrsos* (0–10 cm) because it is the soil-moisture variable consistently archived across all four ESMs. We will state this, note that the 30-year mean of a difference field is far more stable than instantaneous surface moisture, and add total-column soil moisture (*mrso*) for comparison in the Supplement.

– Line 132: Comparing the mean of several UK model runs with individual models is likely unfair as e.g. aspects as internal climate variability are merged out. To enhance comparability, we suggest touse an individual ensemble member, and compare the results with that of the other ensemble members to estimate the relevance of them.

We agree and will use a single UKESM1 member (r4i1p1f2) throughout, for consistency with the single-member treatment of the other three ESMs and because the land-cover diagnostics (*fracLut*) are archived only for that member. To demonstrate that internal variability does not drive our conclusions, we will report the agreement between r4i1p1f2 and the r1–r4 ensemble mean (spatial correlations ≈0.75–0.99 across variables; ≈0.78 for BGWS) and add the inter-member spread in the Supplement.

– Line 144,317, 461: Explain more what you mean with "realized land cover response" and how the same prescribed land use (change) would lead to e.g. different tree cover fractions in different models

We will define "realised land-cover response" (the cover fractions a model actually simulates after the LUH2 transitions are mapped onto its PFTs) and explain that, because models map LUH2 classes onto different PFTs with different tree-cover definitions and — in UKESM1 — interactive competition, identical forcing produces different tree-cover fractions (Loughran et al., 2023).

– Line 145: Why are some LULC types given as fractions and some as tile fractions? Is there a difference between both?

We will add a clarification to Sect. 2.2. *treeFrac* is a vegetation-cover diagnostic — the percentage of the grid cell covered by trees as realised by the model — whereas *fracLut* is a land-use accounting diagnostic reporting the fractional area of each model land-use tile (primary and secondary natural vegetation, cropland, pasture). They thus describe different quantities: realised woody cover versus the partition of grid-cell area among land-use categories, irrespective of the vegetation within them.

- In sec 2.3 maybe give an example between which values BGWS globally varies so that readers can later better understand ΔBGWS[%]

  We will add the typical global range of BGWS so readers can contextualise the magnitude of ΔBGWS [ppts].

- Sec 2.4: Maybe use a heading that tells about the content rather than the method you are explaining

  We will rename the heading accordingly.

- Line 210: Do you observe less forest and more cropland somewhere else than in China? If yes, does the explanation based on bioenergy crops hold there as well?

  We will clarify that this refers to the SSP1 − SSP3 difference, and that under SSP1 cropland expands for second-generation bioenergy in several high-productivity regions. We will name the other regions where SSP1 cropland fractions exceed SSP3 and discuss whether the bioenergy explanation plausibly applies there.

- Line 219-221: Is this already interpretation/discussion? If so, maybe add a reference here

  We will support this interpretation (crop PFTs having higher photosynthetic capacity per unit leaf area) with an appropriate reference.

- Line 222: Check whether it is a hyphen or a minus between GPP and LAI

  We will rephrase this as "the relationship between ΔGPP and ΔLAI" to avoid ambiguity.

- Fig 3: caption: missing 'in' before %-sign; 'In the MMM, regions with low ensemble…' add 'with dots' in the end

  We will correct the caption (add "in" before the %-sign; state that low-agreement regions are indicated with dots).

- Lines 230-234: This is informative content which, however, should be discussed later in this subsection.

  We will move this content to a later point in Sect. 3.2.

- Line 239: signal at the coast of Africa varies quite a lot between models

  We agree the west-coast-of-Africa BGWS signal shows considerable inter-model spread. We will present it as a tendency subject to high inter-model uncertainty rather than as a robust MMM feature, and ensure the low ensemble agreement in this region is clearly marked and discussed (Major comment 1).

– Line 242: does ‘not consistently’ refer to across space or across models?

We will clarify that "not consistently positive" here refers to spatial inconsistency (across grid cells), not across models.

– Line 250-256: Link supplementary figures of changes in P, ET, R

We will add the relevant cross-references to the supplementary ΔP, ΔET and ΔR figures here.

– Check supplementary figures to remove data over the ocean

For ΔET and ΔP we deliberately retained ocean values to keep the global pattern visible and to show potential non-local ET/precipitation responses to LULCC. We will state this rationale in the captions.

– Line 263-275: Maybe add boxes around the example regions you are referring to in the text to make it easier for readers to follow across the different figures

We will add boxes outlining the example regions (eastern US, southwestern Canada, western Amazon, etc.) on the relevant maps to help readers track them across figures.

– Line 275: through a shrinking precipitation denominator - does this refer to eq. 3?

Yes, we will make explicit that this refers to the precipitation denominator P in Eq. (3).

– Line 280: ‘suggesting a tighter local ET-precipitation co-variability’ specify in which of the previously mentioned regions

We will specify that this refers to the eastern US and southwestern Canada (in contrast to the western Amazon).

– Line 285: ‘Similarly’ - to what?

We will clarify that "similarly" refers back to the finding that strong positive LAI/GPP responses do not always yield larger green water shares — i.e. here, negative LAI/GPP responses do not always yield larger blue water shares.

– Line 400: unclear what ‘direction of vegetation change (loss versus gain)’; forest and crops are both vegetation

We will clarify that we mean loss vs. gain in LAI and GPP (vegetation greening vs. browning), not forest-vs-crop land-cover type.

– Lines 455 & 458: The difference between the first and second source of uncertainty is unclear.

We will sharpen the distinction: the *first* source is the translation of LUH2 transitions into a model’s internal land-cover map (how much cover actually changes); the *second* is how those cover fractions are mapped onto PFTs and whether

PFTs compete interactively (which PFTs change, and how). We will rephrase so the two are clearly separated.

References

Hoek van Dijke, A. J., Herold, M., Mallick, K., Benedict, I., Machwitz, M., Schlerf, M., Pranindita, A., Theeuwen, J. J. E., Bastin, J.-F., and Teuling, A. J.: Shifts in regional water availability due to global tree restoration, Nature Geoscience, 15, 363–368, https://doi.org/10.1038/s41561-022-00935-0, 2022.

IPCC: Annex II: Glossary, in: Climate Change 2022: Impacts, Adaptation and Vulnerability. Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, pp. 2897–2930, Cambridge University Press, Cambridge, UK and New York, NY, USA, https://doi.org/10.1017/9781009325844.029, 2022.

Loughran, T. F., Ziehn, T., Law, R., Canadell, J. G., Pongratz, J., Liddicoat, S., Hajima, T., Ito, A., Lawrence, D. M., and Arora, V. K.: Limited Mitigation Potential of Forestation Under a High Emissions Scenario: Results From Multi-Model and Single Model Ensembles, Journal of Geophysical Research: Biogeosciences, 128, e2023JG007 605, https://doi.org/10.1029/2023JG007605, 2023.

Schleussner, C.-F., Lissner, T. K., Fischer, E. M., Wohland, J., Perrette, M., Golly, A., Rogelj, J., Childers, K., Schewe, J., Frieler, K., Mengel, M., Hare, W., and Schaeffer, M.: Differential climate impacts for policy-relevant limits to global warming: the case of 1.5°C and 2°C, Earth System Dynamics, 7, 327–351, https://doi.org/10.5194/esd-7-327-2016, 2016.