Review Report

ESD_2026-2620: **"Global and Regional Hydroclimatic Responses to Alternative Global Reforestation Pathways"** by Mortezapour et al.

Overview

The submitted manuscript investigates the long-term hydroclimatic impacts of two global reforestation scenarios using the CanESM5.1 Earth system model. A major strength of the study is its extension beyond the 21st century and its comparison of two reforestation pathways with different spatial distributions. The manuscript, however, currently makes several interpretations that go beyond what can be supported by the experimental design. In particular, the experiments do not isolate the effect of reforestation location or cleanly separate local from remotely mediated responses, yet these distinctions form an important part of the interpretation. Some mechanistic claims involving the Hadley circulation, AMOC, and ITCZ are also stronger than the diagnostics presented can establish. More generally, the framing is broader than the simulations themselves, which assess the biogeophysical response to prescribed land-cover changes in a single model with atmospheric CO_2 held fixed across experiments. Below are some general and specific comments in this regard.

General Comments

1. **Introduction:**

   The Introduction currently motivates a broader question than the experiment actually answers. It moves among CDR, total climate effects of forests, structural model uncertainty, spatial dependence, water availability, and reforestation strategy, whereas the experiment is more specifically a single-model assessment of the long-term biogeophysical hydroclimatic response to two prescribed global spatial patterns of forest-cover increase under SSP1-2.6. Tightening the Introduction around that narrower contribution would resolve several of the issues at once.

   The later analysis places considerable emphasis on large-scale ocean circulation, including the AMOC, but the Introduction provides little motivation for why reforestation might generate an ocean-circulation response or why this is important for the century-scale hydroclimatic adjustment being investigated. If AMOC changes form an important part of the proposed mechanism, the relevant background and motivation should be introduced here.

   Please define how the terms "reforestation," "afforestation," "forestation," and "forest expansion" are being used. These terms appear interchangeably in the literature review, although they can refer to different land-cover transitions. It would be useful to distinguish the manuscript's own reforestation scenarios from afforestation experiments cited as supporting literature.

2. **Methods:** The experimental design does not fully support attribution of differences between the two pathways specifically to the *location* of reforestation. The sustainable and reversed experiments represent two compound global land-cover perturbations in which many regions are altered simultaneously. They also differ slightly in total reforested area (6.3 versus 6.6 Mkm²), and the manuscript does not present a controlled regional perturbation or formal decomposition that isolates the effect of reforestation location. Although the experiments can demonstrate that two alternative spatial configurations of global reforestation produce different hydroclimatic responses, they cannot determine which particular regional perturbations cause those differences or separate local from remotely mediated responses within reforested areas. A direct statistical comparison between the sustainable and reversed experiments would also be needed if differences between the two pathways are a central result; showing that each differs from the baseline is not equivalent to showing that they differ from each other. The objectives and subsequent attribution should therefore be framed in terms of sensitivity to alternative spatial patterns of reforestation rather than causal isolation of the effect of location.

3. **Results:** Throughout the Results, the manuscript distinguishes between "local" and "remote" responses. Responses outside grid cells where land cover is altered can reasonably be described as non-local/remote responses to reforestation elsewhere. However, responses within reforested grid cells cannot be attributed exclusively to local land-cover change, because these grid cells are also subject to circulation and other remotely mediated responses generated by land-cover changes elsewhere. I suggest that the authors define their spatial criterion for "remote" explicitly and replace "local response" with "response over reforested areas" unless a formal local/remote decomposition is performed.

4. **Discussion:** The Discussion frequently moves from simulated associations to causal attribution more strongly than is supported by the experimental design and diagnostics. In particular, the manuscript states that geographic location is the "main cause" of the differences between the reforestation pathways and interprets changes in the Hadley circulation, AMOC, ocean heat transport, and ITCZ as drivers of the long-term hydroclimatic response. The simulations demonstrate that this circulation and hydroclimatic anomalies occur together, but they do not provide a formal decomposition establishing that the circulation anomalies cause the hydrological response, nor that their effects exceed the direct land-surface response. Similarly, the two global reforestation experiments cannot establish that geographic location is the dominant cause of their differences. I recommend consistently distinguishing between responses demonstrated by the simulations, mechanisms that are consistent with those responses, and causal mechanisms that have actually been isolated by the experimental design. Statements such as "confirm," "main cause," and "driven primarily by" should either be supported by additional attribution analyses or replaced by more cautious language such as "consistent with," "associated with," or "suggests."

5. **Figures**

   i. **Representation of ensemble uncertainty:** The manuscript uses a five-member ensemble and emphasizes the robustness/significance of the simulated responses, but uncertainty is not represented consistently across the figures. Figures 2–3 show statistical-significance stippling, whereas the zonal-mean curves and the regional summaries in Figures 4–5 generally do not show ensemble spread or uncertainty. Figure 6 contains error bars, but the caption does not define what they represent. I suggest showing an appropriate measure of ensemble variability/uncertainty for the regional and zonal diagnostics and defining it consistently in the captions. This is particularly important where differences between the sustainable and reversed scenarios are interpreted as meaningful.

   ii. **Improve readability and make captions self-contained:** Several of the figures contain many panels, relatively small labels, and information that has to be inferred from the main text. I suggest labelling individual panels consistently (a, b, c, …), increasing the size of axis labels/legends where possible, and ensuring that each caption states clearly: (i) whether the quantity is an annual or seasonal mean, (ii) the averaging period, (iii) the experiment/reference used to calculate the anomaly, (iv) the units, (v) any spatial mask or tree-cover threshold applied, and (vi) the meaning of stippling, contours, shading, and error bars. This would substantially improve the figures' ability to stand alone.

   iii. **The visual presentation should clearly distinguish between the imposed forcing and the diagnosed response:** Since one of the main interpretations of the manuscript concerns responses occurring within versus outside reforested areas, it is important that the reader can readily identify where the prescribed land-cover perturbation occurs. Figures 2–3 use the >10% tree-cover contour for this purpose, but the distinction is visually subtle, and the threshold itself is not physically self-evident. Please ensure that the reforestation mask is clearly visible and consistently defined across figures and explain why >10% is used. This would also help avoid visually implying that anomalies within the contour are necessarily "local" responses.

Specific comments and/or recommendations

Abstract

1. Change from CO2 to CO_2

Introduction

2. **Line 30:** Already define the abbreviation "CDR" very close to the terms being defined rather than much later. "… biodiversity, land stewardship, and climate resilience in addition to its well-established role in carbon dioxide removal **(CDR)** …"

3. **Line 31:** The phrase "carbon role" is unclear. If the authors are referring to the role of forests in climate mitigation through carbon sequestration and atmospheric CO_2 removal, please state this more explicitly (e.g., "Beyond their role in carbon sequestration…").

4. **Line 35:** The statement that the hydroclimatic consequences of reforestation "remain less certain and appear to be region- and scale-dependent" is presented as an important knowledge gap. However, regional and scale dependence also appears to be one of the principal conclusions of the present study. It is therefore unclear what aspect of this uncertainty the current study is intended to reduce. Please define the knowledge gap more specifically and explain how the experimental design addresses it. For example, is the novelty primarily the extension beyond the 21st century, the comparison of two alternative global reforestation pathways, or the characterization of long-term regional responses?

5. **Lines 35–37:** The statement that reforestation has "disproportionately bigger impacts on streamflow compared to deforestation" would benefit from greater precision. Does this refer to the hydrological sensitivity of streamflow per unit forest-cover change? If so, please state this explicitly rather than referring broadly to "bigger impacts." Also, please check the citation and corresponding reference entry. The doi "10.1016/j.foreco.2022.120718" points to **Hou et al. (2023)** not **Li et al. (2022)**. Suggest checking the reference list all over.

6. **Lines 38–63:** The literature synthesis alternates between suggesting that a "consistent picture" has emerged and emphasizing substantial model and regional uncertainty. I suggest distinguishing more clearly between mechanisms that are relatively well established (e.g., changes in ET, albedo, and roughness) and the sign/magnitude of the resulting hydroclimatic response, which appears to remain uncertain.

7. **Lines 46-49:** The sentence beginning "These findings demonstrate…" is difficult to follow because several mechanisms with different, and in some cases opposing, effects are combined in a single sentence. In particular, the relationships among increased canopy water demand, reduced albedo, increased ET, cloud feedbacks, drying, and cooling are not clearly expressed. Please restructure this sentence to distinguish the individual mechanisms and their respective hydroclimatic effects.

8. **Lines 55–57:** Please reconsider the term "net precipitation" for P–ET. Since precipitation itself may increase while P–ET decreases, this terminology can be confusing.

9. **Lines 60–63:** The discussion of structural uncertainty concludes that there is a need to better understand "model sensitivity." However, the present study uses only CanESM5.1. An initial-condition ensemble can characterize internal variability but cannot assess sensitivity to model formulation or structural uncertainty. Therefore, either this stated gap should be reframed to match what the present study can address, or the authors should explain more clearly how their analysis contributes to the issue of model sensitivity.

10. **Lines 69–73:** In addition to being partly repetitive, this statement appears internally inconsistent. Forests are first described as "improving infiltration" and then as facilitating "slower infiltration." Is the latter intended to refer to slower runoff or delayed drainage? Please clarify. More generally, statements that forest absence necessarily causes greater quick-flow losses and reduced groundwater replenishment seem overly general given the strong dependence of these responses on climate, soils, forest type, and antecedent conditions.

11. **Line 72:** The statement beginning "Forested landscapes, in contrast…" appears redundant with the preceding discussion of infiltration, overland runoff, and groundwater replenishment. Consider deleting or consolidating these sentences.

12. **Line 78:** The statement that "afforestation-based CO_2 removal scales linearly with carbon uptake" is unclear. CO_2 removal and carbon uptake are closely related quantities, so it is not evident what two variables are being compared here or what "scales linearly" means in this context. Please clarify the intended relationship. If the authors are referring to the approximately linear scaling of carbon-cycle responses with the magnitude of afforestation/reforestation deployment reported by Moustakis et al. (2025), this should be stated more precisely.

13. **Lines 81–83:** The lack of information beyond the 21st century appears to be an important motivation for extending the simulations to 2200, but this gap is not sufficiently established in the preceding literature review. Please summarize the temporal extent of relevant previous studies and explain why post-2100 adjustment is expected to provide new insight. This would make the rationale for the long simulations much stronger. The Introduction also identifies uncertainty "across contrasting land-cover strategies" as a key gap, but the preceding literature review does not clearly establish this gap. It would help to discuss what is already known from studies comparing alternative spatial patterns of reforestation and what remains unresolved.

14. **Lines 84–85:** The "reversed" scenario is described here as restoring "all of the forest that has been lost since 1850." However, the Methods appear to define the scenario using cropland expansion and restoration of forest on land converted to cropland after 1850. Please reconcile these descriptions. If only cropland-driven historical forest loss is reversed, "all of the forest" would seem too broad.

15. **Lines 86–87:** I question the use of "water availability" as a synonym for P–ET. P–ET is an informative climatic water-balance metric, but it does not in itself represent water availability, which also depends on runoff, soil-water storage, groundwater, and temporal variability. This distinction is particularly relevant here because runoff and soil moisture are analyzed separately in the manuscript. I suggest using a more precise term such as "climatic water balance" for P–ET or explaining and justifying why it is being used as a proxy for water availability.

16. **Lines 83–89:** The scope of the simulated climate response should be stated more explicitly in the Introduction. Because atmospheric CO_2 concentrations are prescribed identically across the experiments, the study isolates the biogeophysical effects of reforestation and does not simulate the climatic consequences of additional CO_2 removal/carbon sequestration. Given the extensive motivation in terms of CDR, this distinction is important for interpreting what is meant by the "climate impacts of reforestation." I suggest making this clear when introducing the experimental objectives.

17. **Line 89:** The stated objective "to determine how the location of reforestation influences hydroclimatic outcomes" appears stronger than what can be established from the experimental design. The two reforestation scenarios differ in their global spatial patterns, meaning that the simulations can demonstrate whether two alternative reforestation configurations produce different hydroclimatic responses. However, they do not isolate the causal effect of reforestation location. Doing so would require controlled experiments in which reforestation is varied independently in individual regions (or

otherwise systematically varied) while reforestation elsewhere is held fixed (e.g., Amazon, is reforested alone while other regions are held fixed). I suggest reframing the objective in terms of assessing the sensitivity of hydroclimatic responses to alternative spatial patterns of global reforestation, rather than determining how reforestation location itself influences those responses.

18. **Lines 92–103:** Please specify the exact CanESM5.1 model version/configuration used, including the physics variant (e.g., p1 or p2) and, if possible, the code version/tag. This is important for reproducibility because different CanESM5.1 variants have different model characteristics. The model description should also cite the publication specifically documenting CanESM5.1 (Sigmond et al., 2023). A table in the Supplemental materials could be helpful for this.

Methods

19. **Line 99:** Please define the acronym at its first mention: "nine plant functional types (PFTs).

20. **Lines 105–120:** More information is needed on how the two reforestation scenarios were translated into the prescribed PFT distributions used by CanESM5.1. For the sustainable scenario, how was the reforestation opportunity map of Griscom et al. (2017) remapped to the CanESM grid, and how was the target forest fraction in each grid cell determined? For both scenarios, which existing PFT fractions were reduced to accommodate additional tree cover, and how were the newly added tree PFTs assigned (e.g., broadleaf/needleleaf, evergreen/deciduous)? These choices are important because the resulting albedo, roughness, rooting characteristics, and evapotranspiration depend on the forest PFT being introduced. Please provide sufficient detail to reproduce the prescribed land-cover perturbations.

21. **Lines 110-117:** The manuscript states that tree cover is prescribed to increase linearly between 2015 and 2070. Please provide sufficient detail on how this was implemented. For example, was the forest fraction in each grid cell increased by a constant annual increment from its 2015 value to its prescribed 2070 target? What land-cover/PFT fractions were reduced correspondingly, and how were the added forest PFTs assigned? It would also be useful to report the corresponding global annual rate of forest-area increase for each scenario. A description analogous to that provided for the deforest-glob experiment "*Within 50 years, 20 Mkm^2 (million square kilometres) of forest area is removed in a linearly increasing manner, at a rate of 400 000 km2 yr^{-1}.*" in Lawrence et al. (2016; 10.5194/gmd-9-2973-2016), and Boysen et al. (2020; 10.5194/bg-17-5615-2020) would improve reproducibility.

22. **Line 116:** Please replace "In both cases" with "In both reforestation scenarios" to avoid ambiguity.

23. **Lines 117–120:** Please provide the source and definition for the reported global forest areas of 75 Mkm² in 1850 and 68 Mkm² in 2015. Are these observational estimates, reconstructed LUH tree-PFT areas, or CanESM land-cover fractions? The manuscript also alternates between "tree cover" and "forest area"; please clarify whether these quantities are equivalent in the model. Referring to the 1850 value as "observed" may also require reconsideration given that historical land cover is reconstructed rather than directly observed.

24. **Lines 117–124 / Figure 1**: Figure 1 contains negative tree-cover changes although both experiments are described as reforestation scenarios and the caption refers to "percentage increase" in forest cover. Please explain the origin of these forest-cover decreases. Also clarify whether the reported 6.3 and 6.6 Mkm² changes represent gross forest gain or net global forest-area change after accounting for decreases.

25. Lines 119–120 and 139–150: If a central objective is to determine whether the different spatial distributions of reforestation produce different hydroclimatic responses, I would expect a direct comparison between the sustainable and reversed experiments, including an assessment of whether their differences are statistically significant. Demonstrating that one experiment differs significantly from the baseline while another does not, is not equivalent to demonstrating that the two experiments differ significantly from one another. Please consider including the direct contrast $R_{reversed} - R_{sustainable}$.

26. Figure 1/ Lines 123 – 124: Please clarify exactly how the plotted tree-cover change is calculated. Does "percentage increase" denote the absolute change in grid-cell forest fraction in percentage points, $100(F_{2070} - F_{2015})$, or the relative percentage change, $100(F_{2070} - F_{2015})/F_{2015}$? This distinction is important, particularly because the figure also appears to contain negative values despite being described as an "increase." Please also explain why 2070 was selected as the endpoint of reforestation and state explicitly whether land cover is held fixed at its 2070 configuration thereafter until 2200.

27. Lines 126–132: Please clarify and justify the choice of a fixed-2015 land-cover baseline. Although atmospheric and other external forcings follow SSP1-2.6, the baseline does not appear to include the evolving land-use trajectory associated with SSP1-2.6. It therefore represents an idealized no-future-land-cover-change counterfactual rather than a standard SSP1-2.6 simulation. This distinction should be made explicit, particularly when describing the experiments as being conducted "under SSP1-2.6."

28. Line 127 and throughout: Please use consistent capitalization and formatting for the experiment/scenario names (*historical, baseline, sustainable*, and *reversed*) throughout the manuscript. Also, either call them sustainable scenario or sustainable pathway. It is currently used interchangeably.

29. Lines 128–131: Because the post-2100 response is a central focus of the study, please describe the forcing protocol from 2100 to 2200 in greater detail. How are CO_2 concentrations, non-CO_2 greenhouse gases, aerosols, and other relevant external forcings extended beyond 2100? Please provide the specific extension protocol/dataset and clarify whether any forcings are held constant after 2100. Likewise, please state explicitly whether the prescribed land-cover distributions remain fixed at their 2070 values through 2200.

30. Lines 132–136: Please provide further details on the generation of the five ensemble members. Which ocean-state variable(s) were perturbed, what was the magnitude/form of the perturbation, and were identical perturbations used to generate corresponding members of the baseline, sustainable, and reversed experiments? The latter is particularly important because the Wilcoxon signed-rank test assumes paired samples. If member (*i*) of each experiment does not share the same perturbed initial state, the use of a paired test would require justification.

31. Line 141 and throughout: Please specify the temperature variable used consistently. In CMIP6 terminology, surface temperature (*ts*) and near-surface air temperature (*tas*) are distinct quantities. Figure 2 refers to 2-m air temperature, suggesting that *tas* is being analyzed. First check this to confirm that *tas* for CanESM5.1 is at 2m. And if so, please use "near-surface air temperature" consistently rather than "surface temperature." More generally, providing the corresponding CMIP6 variable names for the analyzed fields would improve reproducibility.

32. Lines 141 – 143: The distinction between directly analyzed model variables and derived diagnostics should be made clearer. P−ET is a derived quantity rather than an individual model variable, while the Hadley circulation and AMOC are circulation diagnostics. Please specify how P−ET is calculated, including the sign convention for ET, and describe how the Hadley circulation and AMOC metrics are derived (e.g., streamfunction calculation, latitude/pressure or depth range, and definition of circulation strength).

33. **Lines 143–145:** Please define the seven regional domains quantitatively, preferably by providing their latitude–longitude bounds in a table. Also specify how regional and zonal means are calculated: are they area-weighted by grid-cell area, restricted to land grid cells, restricted to reforested grid cells, or calculated over the entire regional domain? These choices can materially affect regional hydroclimatic averages.

34. **Lines 146–147:** Please provide considerably more detail on the statistical significance testing. What constitutes the sample supplied to the Wilcoxon signed-rank test: the five ensemble-member differences, annual values within each 21-year climatological period, or some combination of ensemble members and years? If the five ensemble members constitute the sample ($n = 5$), please explain how $p < 0.05$ is obtained, since an exact two-sided Wilcoxon signed-rank test with five non-zero paired observations has a minimum attainable p-value of 0.0625. If an asymptotic approximation or a different sample definition is used, this should be stated and justified. In addition, please clarify whether corresponding ensemble members across the baseline and reforestation experiments are paired through identical initial-condition perturbations, as required for a signed-rank test.

35. **Lines 148–150:** Please justify using 2015–2035 as the "2025 climatology" for temporal anomalies. Reforestation is already underway throughout this period, so it is not a stationary pre-reforestation reference state. If temporal changes are being interpreted as responses since the onset of reforestation, a pre-2015 reference period may provide a cleaner baseline. If the present choice is retained, its implications should be explained.

36. **Lines 149 – 150:** The statement that "most differences are evaluated relative to the baseline simulation" is rather ambiguous. Please specify which analyses are not evaluated relative to the baseline and clearly distinguish between (i) reforestation-minus-baseline differences at a given time and (ii) temporal changes relative to the 2015/2025 reference period. See later comment on 2025.

37. **Line 150:** Please define precisely what is meant by the "2025 climatology." Is this the 2015–2035 mean defined above, or the value for the individual year 2025? Also specify whether anomalies are calculated relative to the baseline 2025 climatology or separately relative to each experiment's own 2025 climatology. Finally, "normalized by reforested area (Mkm²) where relevant" requires further explanation: please identify which analyses are normalized, whether global or regional reforested area is used, and provide the calculation explicitly.

38. **Line 150:** Please also justify the normalization by reforested area. Dividing a hydroclimatic response by Mkm² implicitly facilitates interpretation in terms of response per unit reforested area, but such scaling may not be appropriate if the response to forest-cover change is nonlinear and strongly location-dependent, as argued elsewhere in the manuscript. Please clarify the purpose of the normalization and whether the conclusions depend on it.

Results

39. **Line 154 and throughout:** As noted earlier, please clarify whether the analyzed variable is surface temperature (ts) or near-surface air temperature (tas). Figure 2 indicates that 2-m air temperature is used, in which case "near-surface air temperature" should be used consistently.

40. **Line 155:** The current sentence could be read as implying that year-2015 land cover is held fixed in all experiments. My understanding is that this applies only to the baseline experiment. Please restructure the sentence to make this distinction explicit.

41. Lines 156–157: “Relative to the fixed land-cover baseline” is repetitive given the preceding sentence and could be removed.

42. Line 157: I suggest avoiding the causal wording “increased tree cover yields widespread cooling.” The simulated anomaly at a reforested grid cell represents the combined response to the local land-cover perturbation and any remotely mediated atmospheric/oceanic response to reforestation elsewhere. The present experimental design does not separate these contributions. Consider instead stating that “cooling occurs over many reforested tropical regions.”

43. Line 165 and throughout: Please use consistent terminology and capitalization for the two experiments (e.g., “sustainable scenario” and “reversed scenario” throughout).

44. Line 165: Referring to these anomalies as “local responses” is stronger than the experimental design allows. Responses over reforested grid cells may contain both locally forced and remotely mediated components. I suggest referring instead to “responses over reforested regions,” unless a method is provided to isolate the local contribution.

45. Lines 169–170: Please define what is meant by “remote effects.” If this refers to anomalies in grid cells where tree cover itself was not altered, that spatial definition should be stated explicitly. Conversely, anomalies within reforested grid cells should not necessarily be described as purely “local,” because they may contain both local and remotely mediated components.

46. Lines 187–191: If “remote effects” are defined as responses outside regions of prescribed tree-cover change, please state this explicitly. However, attribution of these responses specifically to “adjustments in circulation and energy transport” requires supporting diagnostics. Spatial separation from reforestation can establish that the response is non-local but does not by itself identify the mechanism responsible.

47. Line 197: Section 3.2 subheadings: Given the attribution issue noted above, I suggest replacing “Local responses in reforested regions” with “Responses over reforested regions” and defining “Remote and non-local responses” explicitly as responses outside areas of prescribed forest-cover change.

48. Line 198: Consider replacing “promotes higher evapotranspiration” with the more neutral “results in higher evapotranspiration” or “is associated with higher evapotranspiration.”

49. Line 205 and throughout: Please use a consistent convention for numerical ranges. For example, write positive increases as “+0.1 to +0.3 mm day^{-1}” and negative anomalies as “−0.3 to −0.2 mm day^{-1},” or alternatively “decreases of 0.2–0.3 mm day^{-1}.” Avoid constructions such as “+0.1–0.3,” where the sign applies ambiguously to the range.

50. Lines 205–208: The statement that “most of the added rainfall is recycled through evapotranspiration” does not appear to be demonstrated by the analyses presented. Co-located increases in ET and precipitation are consistent with enhanced moisture recycling, but they do not establish either the source of the additional precipitation or the fraction attributable to recycled terrestrial moisture. Demonstrating the latter would require a moisture-recycling or moisture-source diagnostic. Please either provide supporting analysis or substantially soften this attribution.

51. Line 213: Since the increase in near-surface specific humidity is used as supporting evidence for the proposed moisture-recycling mechanism, please show this diagnostic, at least in the Supplement, rather than referring to it as “not shown.”

52. **Line 217 and figure captions throughout:** Please specify explicitly whether reported maps and numerical values represent annual means or individual seasonal means. Although the Methods state that both annual and seasonal analyses are performed, this is not always clear in the Results and figure captions.

53. **Lines 217–218:** The sentence is ambiguous regarding the association between region and season. Are the largest reductions in South America associated with DJF and those in West Africa with MAM, or do both regions exhibit reductions during both seasons? Please rewrite to make the regional and seasonal associations explicit.

54. **Line 220:** "Greater canopy water demand" is vague and appears mechanistically imprecise. If the authors mean increased evapotranspirative losses associated with transpiration and interception evaporation, please state this explicitly. Attributing the runoff decrease specifically to canopy demand also requires supporting evidence.

55. **Figure 3/ Lines 224 – 228:** Please state explicitly in the caption whether the plotted fields are annual means for 2090–2110. I also recommend labelling the individual panels (a, b, c, …), which would make it considerably easier to refer to specific variables/scenarios in the text.

56. **Lines 253–257:** The terrestrial water-budget equation and definitions of its components are methodological information and should be introduced in Section 2 rather than for the first time in the Results. Please define S, the units and averaging procedure used for dS/dt, and how the individual model variables entering the budget are calculated.

57. **Lines 257–260:** Please clarify how the >10% tree-cover threshold is used. Are the hydrological anomalies themselves calculated only over grid cells exceeding this threshold, or are anomalies calculated over the entire regional domain and subsequently divided by the area of grid cells exceeding 10% tree-cover increase? Please also justify the choice of the 10% threshold and indicate whether the results are sensitive to it.

58. **Line 265:** Please distinguish clearly between $P - ET$and the complete terrestrial water budget. The equation above defines the storage tendency as $dS/dt = P - ET - R$, whereas the discussion here focuses on $P - ET$. P−ET is not equivalent to the storage tendency because runoff has not been removed. The purpose of using P−ET as a diagnostic should therefore be stated explicitly.

59. **Lines 261–270:** Much of this paragraph repeats precipitation and ET responses already described in Section 3.2. Since this section is explicitly framed as a regional water-budget analysis, I suggest focusing this section on how the individual terms combine to determine the net regional budget rather than repeating the individual flux anomalies.

60. **Section 3.3/Lines 277 – 287:** Although this section is presented as a "Regional Surface Water Budget" analysis and introduces $dS/dt = P - ET - R$, the results primarily show P, ET, R, P−ET, and soil-moisture anomalies separately. The manuscript does not appear to calculate or show dS/dt, nor does it demonstrate closure of the stated water budget. A soil-moisture anomaly is also not equivalent to a storage tendency. Please either provide an explicit water-budget closure analysis, including the storage tendency/residual, or revise the framing of this section as an analysis of regional hydrological components.

61. **Line 270 / Figure 4:** Please check the units used for the area normalization. The caption states that the quantities are normalized by reforested area in km², whereas the axes and main text report changes per Mkm² of reforested area. Please use a consistent unit throughout. The "Drier" and "Wetter" labels may be misleading because the panel combines different hydrological fluxes whose signs do not have a common interpretation; for example, positive ET does not necessarily indicate wetter conditions. I

suggest removing these labels or restricting such terminology to a diagnostic for which the sign can clearly be interpreted in terms of wetting/drying.

62. Lines 284–286: The approximately 4-m CLASS soil-column depth is methodological information needed to interpret the soil-moisture variable and should be defined in the Methods. The figure caption could then briefly remind the reader that the reported values represent vertically integrated soil moisture.

63. Line 294: Please state the direction of the anomalies rather than referring generically to "modest but sustained anomalies," e.g., "sustained increases in evapotranspiration and decreases in runoff…"

64. Figure 5: Please clarify the calculation represented by each curve. The caption currently states that the changes are calculated as the "difference between 2100 and 2190," whereas the figure displays separate curves labelled 2100 and 2190. My understanding is that each curve may instead represent the reforestation-minus-baseline anomaly calculated independently for the corresponding climatological period. If so, please revise the caption accordingly. This distinction is important because the section is intended to demonstrate how the reforestation response evolves after 2100.

65. Line 307 and throughout: Please standardize the terminology and capitalization used for experiment names. The manuscript alternates among "Sustainable scenario," "Sustainable pathway," "Reversed scenario," "Reversed case," etc. Choose one convention and apply it consistently.

66. Figure 6 / Line 326: Please change "2m air temperature" to "near-surface (2-m) air temperature," consistent with the terminology used elsewhere. The caption and y-axis also refer to a "2025 baseline," whereas elsewhere "baseline" denotes the experiment with land cover fixed at its 2015 state. Please clearly distinguish between the baseline experiment (fixed 2015 land cover) and the 2025 reference climatology used to calculate temporal percentage changes and specify exactly how the latter is calculated.

67. Figure 6: I have some difficulty interpreting the calculation shown here. If the values at 2100 and 2190 are expressed relative to the 2025 baseline climatology, they will contain both the response to reforestation and the background temporal climate response to SSP1-2.6. This seems inconsistent with the rest of the manuscript, where the effect of reforestation is isolated by comparing each reforestation experiment with the baseline experiment at the same time. Please clarify the purpose and calculation of Figure 6 and consider whether same-time reforestation-minus-baseline anomalies would be more appropriate. In addition, expressing near-surface air-temperature changes as percentages is not physically meaningful if temperature is expressed in °C, and percentage changes in P−ET may be unstable where the reference P−ET approaches zero. Absolute anomalies may therefore be preferable.

68. Lines 313–323: The attribution of the post-2100 hydroclimatic evolution primarily to changes in the Hadley circulation and AMOC appears stronger than can be established from the diagnostics presented. Figure S7 demonstrates coincident circulation anomalies, but this alone does not establish that these changes *drive* the precipitation response, nor that their contribution exceeds the direct land-surface response. Please quantify the magnitude and ensemble robustness of the AMOC/Hadley-cell changes and provide stronger evidence linking these changes to the hydroclimatic anomalies. Otherwise, I suggest using more cautious wording such as "consistent with" or "associated with" circulation adjustments. Similarly, the statement that circulation effects "exceed the direct biophysical impacts of land-cover change alone" would require an explicit decomposition that does not appear possible from the current experiments.

Discussion

69. **Line 352:** The attribution of the comparatively weak Southern Hemisphere extratropical temperature response to the "smaller land fraction and more efficient ocean heat uptake" is not directly demonstrated by the analyses presented in this study. In particular, ocean heat uptake is not diagnosed here. Please provide appropriate supporting citation(s) and/or soften the statement to indicate that this is a proposed interpretation rather than a result directly established by the simulations.

70. **Line 359:** Please replace "surface temperature" with "near-surface air temperature," assuming the analyzed variable is tas/2-m air temperature and use this terminology consistently throughout the manuscript.

71. **Line 363:** Please move the supporting citations immediately after the statement that mid- and high-latitude forest responses can be dominated by surface-albedo changes in the radiative balance. This would make clearer which statement the cited studies support.

72. **Line 382 / Section 4.4:** I suggest reconsidering the framing of this section as "Drivers of remote effects." The analyses demonstrate associations between the hydroclimatic response and changes in the Hadley circulation, ITCZ, and AMOC, but do not appear to isolate these circulation changes as causal drivers. In particular, the inferred sequence from AMOC weakening to reduced northward ocean heat transport and a compensating atmospheric/ITCZ response is not directly demonstrated by the experiments. Unless additional attribution diagnostics are provided, more cautious terminology such as "Large-scale circulation responses associated with non-local hydroclimatic changes" would better reflect what can be established from the simulations.

73. **Lines 383–384:** The statement that the results "confirm" that variations in the geographic location of forest expansion are the "main cause" of the different hydrological responses appears too strong for the experimental design. The two scenarios demonstrate that alternative global reforestation configurations are associated with different hydroclimatic responses, but they do not isolate the causal contribution of location or establish that location is the dominant cause of the differences. Other differences between the scenarios, including their precise reforested area and potentially their temporal and PFT distributions, may also contribute. I suggest reframing this conclusion in terms of sensitivity to alternative spatial patterns of reforestation, rather than confirmation of geographic location as the main causal factor.

74. **Line 386:** The phrase "Hadley cell differences" is ambiguous regarding which experiment is being discussed. If this sentence refers to the sustainable scenario, please state this explicitly, e.g., "In the sustainable scenario relative to the baseline, the Hadley-cell anomaly shows a broader ascent footprint into the Southern Hemisphere…" The following sentence can then introduce the corresponding response in the reversed scenario.

75. **Line 399 / Section 4.5 (Limitations):** An important limitation that should be stated explicitly is that atmospheric CO_2 concentrations are prescribed identically across the experiments. Consequently, the study quantifies only the biogeophysical hydroclimatic response to reforestation and does not include the climatic effect of additional carbon sequestration/CO_2 removal. This is particularly important for interpreting the policy discussion: the simulations cannot assess the net climatic effect of reforestation or quantitatively evaluate trade-offs between its carbon-removal benefit and its biogeophysical warming/cooling effects. Please make this scope explicit in the limitations and policy implications.

76. **Line 399 / Section 4.5:** Given the emphasis on regional hydrological responses, the relatively coarse atmospheric resolution of CanESM5.1 (~2.8°) should also be acknowledged as a limitation. This resolution cannot resolve individual catchments, complex topography, or fine-scale land–atmosphere interactions and may limit the interpretation of regional precipitation and runoff responses. This is particularly relevant to the basin-scale and water-resource implications discussed later.