

**Soil moisture-induced changes in land carbon sink projections in CMIP6,**  
<https://doi.org/10.5194/egusphere-2025-4215>

**Response to Anonymous Referee #1**

Dear referee,

Thank you for the constructive and helpful comments on our paper.

In our responses to the specific comments, we hereafter refer to the line numbers of the marked-up manuscript. In addition, we further corrected a few grammatical errors during revision.

Overarching comments:

This paper address water-carbon coupling in models leveraging soil-moisture sensitivity runs through the LFMIP experiments. The topic is important to the modeling community, and the work presented here is thorough and yields many good insights. I congratulate the authors on pulling together what is clearly a lot of work. However, this was a difficult read and left many points of confusion; I suggest there is room to further develop the presentation for clarity, primarily through 1) the inclusion of clear guiding hypotheses/questions that motivate and structure the analysis; 2) more consolidation of uncertainty information in the figures; and 3) clearer organization of information throughout.

Thank you for the overall positive evaluation of our work. Your suggestions are very helpful to clarify the research questions addressed in this study, include additional uncertainty information, and improve the organisation of the presented information. Detailed responses are provided with the specific comments below.

Some specific comments are below.

**L 30:** Please define the sign of land source/sink in this terminology. For example., “The net carbon exchange on large spatial and temporal scales is referred to as Net Biome Production (NBP), where positive values indicate a net flux of carbon from atmosphere to land (sink) and negative values indicate a net flux of carbon from land to atmosphere (source).”

This was implemented as suggested.

The sentence extension “[...], where positive values indicate a net flux of carbon from atmosphere to land (sink) and negative values indicate a net flux of carbon from land to atmosphere (source).” was added in the revised version.

**L 55:** Although later in the paper (L 130) the focus on GPP over respiration is asserted, it is worth mentioning the important and variable impacts of SM on microbial respiration, CO<sub>2</sub>/CH<sub>4</sub> partitioning, etc, as the context is set in the introduction. This is a major part of the hydrological feedback puzzle especially at high latitudes where, in addition, total soil C is highly uncertain.

Thank you for pointing this out. Although we focus on the effect of SM on GPP (and therefore only explained processes related to GPP so far), SM induced respiration changes should be mentioned in this context.

The paragraph was rephrased to include the effects on ecosystem respiration:

“SM limitation impacts both major carbon sink components, GPP and ecosystem respiration. Decreased SM can cause water stress in vegetation due to direct physiological effects of water limitation on GPP and indirect effects via land–atmosphere feedbacks (Green et al., 2019; Liu et al., 2025; Zhou et al., 2019). Direct SM effects arise from photosynthetic activity being constrained by soil water availability. Under dry conditions, plants partially close their stomata to limit water losses through transpiration, which in turn reduces CO<sub>2</sub> uptake and thus limits GPP (Sippel et al., 2018). Indirect SM effects arise from interactions between the land surface and the atmosphere. Low SM suppresses latent heat flux and evapotranspiration, leading to higher surface temperatures and vapour pressure deficit (VPD). Low SM suppresses the latent heat flux through evapotranspiration, causing a shift towards sensible heat flux that increases near-surface air temperature and vapour pressure deficit (VPD). Elevated temperatures and increased VPD enhance atmospheric evaporative demand and promote further stomatal closure, thereby reinforcing the direct limitation of low SM on GPP (Berg et al., 2016; Seneviratne et al., 2010). Low SM can also increase incoming radiation by reducing cloud formation and atmospheric humidity, which enhances photosynthetic activity (Seneviratne et al., 2010, Zscheischler et al., 2015). In many ecosystems globally, reduced SM has a more pronounced negative impact on NBP through GPP than respiration (Gu et al., 2017; Li et al., 2023; Von Buttlar et al., 2018). Nevertheless, ecosystem respiration is also sensitive to water scarcity. Low SM can suppress soil respiration by reducing the availability of soluble carbon substrates to soil microbes, while elevated temperatures can lead to higher soil respiration (Gentine et al., 2016).”

**L66-75:** Please introduce some hypotheses or questions here that can be addressed. This will make the results section far easier to read and greatly improve the presentation.

Thank you for this suggestion, we will reformulate this paragraph so that the motivation of the analysis is clearer. The analysis consists of two parts (section 3.1 and section 3.2) which can be addressed with the following two research questions:

1. Is the SM-induced impact on the land carbon sink (NBP) according to LFMIP (CMIP6) different from that according to GLACE-CMIP5 (section 3.1)?

The results of this analysis reveal that LFMIP and GLACE-CMIP5 indeed show large differences in the magnitude of the SM-induced changes of NBP, but also large intermodel uncertainties within each generation. Therefore, we further explored the origins of these differences, which brought us to the second research question:

2. Within LFMIP (CMIP6), what are the origins of (i) SM-induced changes in land carbon uptake (GPP) (section 3.2.1) and (ii) related intermodel differences (section 3.2.2)?

The paragraph was rephrased to:

“In this study, we assess the impact of SM on the land carbon sink in the latest generation of Earth System Models (ESMs) from the Coupled Model Intercomparison Project Phase 6 (CMIP6) using dedicated experiments of the Land Feedback Intercomparison Project (LFMIP) and compare it with results from GLACE–CMIP5. Furthermore, we conduct a contribution analysis to quantify the extent to which SM–induced changes in land–carbon dynamics emerge through direct linear effects of water limitation on photosynthesis, versus indirect effects mediated by land–atmosphere feedbacks and investigate the origins of intermodel differences. Thus, our analysis constitutes of two parts addressing the following research questions:

1. Is the SM-induced impact on NBP according to LFMIP different from that according to GLACE-CMIP5 with regard to the effect of SM trend and variability (see Sect. 3.1)?
2. Within LFMIP, what are the origins of (i) SM-induced changes in GPP (see Sect. 3.2.1) and (ii) related intermodel differences (see Sect 3.2.2)?

Our analysis aims to further improve the understanding of SM–carbon coupling in state–of–the–art ESMs and its implications for long–term carbon cycle projections. Gaining a deeper understanding of the moderating processes within the land carbon cycle is essential for enhancing projections of atmospheric CO<sub>2</sub> growth rates, which in turn is crucial for accurately estimating the magnitude of future climate change (Arora et al., 2020; Friedlingstein et al., 2025).”

**L 85:** Please provide **a)** some brief acknowledgement that there’s no water budget closure when SM levels are prescribed as such and **b)** brief explanation on why the logic and feedback physics of these sensitivity experiments nonetheless works for the variables of interest.

Thank you for pointing this out. When introducing the experiments in the method section, a disclaimer was added to acknowledge this fact:

“LFMIP consists of three experiments [...]. We note that for the experiment simulations pdLC and rmLC, the hydrological cycle is not closed because by prescribing SM externally, hydrological feedbacks from the atmosphere to SM are suppressed. Consequently, water can artificially be added or lost. However, as our focus lies on the SM-induced impact on the carbon cycle and not on the hydrological cycle itself, all relevant processes, including changes in surface energy partitioning, near-surface air temperature, VPD, and physiological controls on photosynthesis remain physically consistent. Therefore, differences between experiment simulations can be robustly attributed to SM-induced effects on NBP and GPP despite the lack of water budget closure.”

**L125:** It will be clearer if the content of this sentence can be reformatted as equations: “The experiments of LFMIP allow isolating the effect of SM trend and variability, where

$\Delta\text{NBPSM trend} = \Delta\text{NBPrmLC} - \text{pdLC}$  and

$\Delta\text{NBPSM var} = \Delta\text{NBPCT L} - \text{rmLC}$  , as well as the combined effects of SM expressed as

$\Delta\text{NBPSM all} = \Delta\text{NBPSM trend} + \Delta\text{NBPSM var} = \Delta\text{NBPCT L} - \text{pdLC}.$ ”

The paragraph was reformulated as equations as suggested.

**L145:** If focusing on total SM, the methods set up can be simplified for the aid of the reader by limiting description/equations to just the pdLC and CTR experiments.

In the first part of the analysis (section 3.1) variability and trend are analysed separately, so the equations are needed. In the second part focusing on GPP (section 3.2), the analysis is performed for the total effect only. We clarified this in the context of the newly added research questions in the introduction (see comment above) and in the respective sections.

In line 176 we added: “We note that **for this part of the analysis** we focus only on the total SM effect (and not on the individual effects of SM trend and variability).”

**L175:** (related to comment on L145) Yet, if there is substantial discussion of SMtrend and SMvar, the authors should leave these definitions in. But then I would rephrase this to say that the other results are primarily in the supplement but discussed throughout

See answer to L145.

**L184-186:** negative signs are redundant with phrasing: “reducing by...”

Thank you for pointing this out. The negative sing was removed.

**Figure 1:** When suggesting a comparison to Green et al 2019 figure 1, please provide more direction on what comparison should be made. For instance, is Figure 1 panel b meant to be a reproduction (over a different time period) of data in Green et al 2019 figure 1? If so, this is confusing, as the black, blue and pink lines do not have the same decadal-scale dynamics—this paper’s lines reach a minimum in ~2020 while Green et al 2019 lines have an approximately

monotonic increase. What accounts for the difference, how should we understand this comparison?

Thank you for this remark, we understand that it is confusing. The purpose of reproducing the analysis of Green et al. (2019) was to compare their results for GLACE-CMIP5 with our results for LFMIP (CMIP6). However, we could not confirm all their decisions during data processing. Differences in figure appearance may, e.g., stem from differences in time aggregation and plot scaling. To ensure a consistent comparison, we reproduced Figure 1 of Green et al. (2019) to the extent possible (panel b) and created the same plot for LFMIP (panel a). Consequently, a visual comparison with their figure is not necessary and the sentence in the caption “For panels (a) and (b) the colour scheme was adapted from Green et al. (2019) to facilitate comparison with their study” is indeed confusing and will be removed. To acknowledge that we adopted the figure design of Green et al. (2019), we will state “For panels (a) and (b) the figure design was adopted from Green et al. (2019)” to highlight that our study builds on their analysis.

**L 200-210/figure 3 (also applies to figure 2):** This analysis would be enhanced by presentation of uncertainty in the figures as follows: **a)** in the model-specific plots, some indication on where the trends are insignificant with stippling or shading. Without this it is hard to assess in z-score space what is going on in the Sahara—some of the Sahel SM increase in CMCC-ESM2 may be “real” accompanying Sahel greening, but some of the other widespread changes over the desert (e.g. IPSL-CM6A-LR) may be noise. Similarly, **b)** some indication in the ensemble mean plots where there is sign disagreement within the 4 models would greatly enhance the interpretation of the visuals. The IPCC reports have good inspiration on how to make these multi-map plots more useful by conveying this type of uncertainty information.

Thank you for these suggestions, we implemented both suggestions, a) and b), to Figure 2 and 3. As our analysis focuses on two 30-year time periods, namely the baseline (1981-2010) and the future period (2070-2099), we indicated significant changes in the future period relative to the baseline by adding stippling in the model-specific plots. For model mean plots, we added hatching for sign disagreement. We extended the captions to include this information. In addition, we applied the same approach to related Figures in the supplement (Supplementary Figure S5 to S7).

**L 256-265:** This is not adequately interpreted. **A)** It would be helpful to state the hypotheses guiding this analysis. Related to that, **B)** to help us understand a statement such as “the results indicate that about 70–90 % of intermodel difference can be explained by either changes in the direct and indirect SM effects or the sensitivity of GPP to those effects” please list out what the possible contributors are, and how this relates to the guiding hypotheses. When I dig back to Eq4 and line 156, this statement appears to suggest that 70-90% of the effects are accounted for by “everything” which is a trivial conclusion. Please make this clearer for the

reader. **C)** Please correct North South America **D)** 263-265 suggest to put this sentence higher in the paragraph, this is where the real conclusions are.

We apologize for the poorly organized paragraph. It will be edited to clarify the content.

a) The hypothesis guiding this analysis was introduced in the introduction (see previous comment). Additionally, the guiding research questions for this part were repeated at the beginning of Section 3.2:

“With focus on LFMIP, we further analyse the origins of SM-induced changes in land carbon uptake by (i) conducting a contribution analysis to assess the impact of direct and indirect SM effects on GPP (Sect. 3.2.1) and (ii) investigating causes of intermodel differences in SM-induced changes in GPP from direct and indirect SM effects and their respective sensitivities (Sect. 3.2.2). To assess the impact of SM on land carbon uptake under future climate change, we focus on the future period (2070–2099) when impacts of both SM variability and SM trend come into play. The results for GLACE–CMIP5 are displayed in supplementary Figs. S15 to S17.”

b) and d): The paragraph was restructured to enhance clarity:

“We perform a factorial ANOVA to spatially assess contributions of direct and indirect SM effects, as well as of the sensitivity of GPP to those drivers as described 2.4. Overall, our results show that in areas that experience a strong reduction in GPP due to SM, the linear approach of the factorial ANOVA captures 70-90% of intermodel differences, which relate to either differences in the evolution of direct and indirect SM effects or the sensitivity of GPP to these drivers (Supplementary Fig. S14). The remaining 10-30% include non-linear effects, lagged effects (e.g., slow-onset effects after drought), and non-local influences (e.g., due to atmospheric transport).

Disagreement in the sensitivity of GPP to SM contributes most strongly to intermodel differences across large parts of the globe, especially in regions with large contributions to global  $\Delta$ GPP like Northern South America, North America and Europe, where models largely agree on SM drying (Fig 7a.3). While disagreement in the sensitivity to the direct SM effect dominates in the tropics (Fig. 7a.1), disagreement in the sensitivity to T and VPD dominates in mid and high latitudes (Fig. 7a.2). A similar but less pronounced spatial pattern can be seen for the contributions of changes in the direct and indirect SM effects themselves (Fig. 7b).”

For Eq. 4, an explanation of the error term epsilon was added in the method section (line 189): “ $\epsilon$  represents the residual change in GPP which is not captured by the linear contributions of direct and indirect SM effects, including non-linear, non-local and lagged effects.”

c) This was corrected to northern South America.

**Overall discussion:** This discussion has several good insights but more organization of information is needed to make this compelling and easy to digest. For example, I could see a header “Are the models fit for purpose?” and another one “Are the experiments fit for purpose?” to help frame the current discussion of model process richness and drought impact

performance and the inclusion of SM extremes, respectively, alongside some indication in the opening discussion paragraph that the authors intend to explore these ideas? This is just an example to illustrate the suggestion of how to organize the information, the authors can of course approach this how they want.

Thank you for this constructive input. The discussion was reorganised to enhance clarity.

An opening paragraph was added at the beginning of the discussion to better guide the reader:

“In this study we investigate (i) the comparison of the SM-induced impact on NBP simulated by LFMIP to the previous experiment generation, GLACE-CMIP5, with regard to the effect of SM trend and variability (Sect. 4.1) and (ii) origins of SM-induced changes in GPP (Sect. 4.2.1) and related intermodel differences focusing on LFMIP (Sect. 4.2.2). In view of our results, we discuss potential causes of the observed intermodel differences and limitations of our analysis (Sect. 4.3).”

We reorganised the subsections such that they are aligned with the overall structure of our manuscript by organising the second research questions into two subsections (i.e., Sect. 4.2.1 and 4.2.2). Additionally, we discuss limitations and outlook in a sperate section (Sect. 4.3).

**L 276-277:** “The latitudinal NBP of the northern mid–latitudes accounts for about 85 % of global NBP for the CMIP5 MMM.” Can this be rephrased? I don’t know what is meant here.

Thank you for pointing this out. It was rephrased to “The NBP in the northern mid-latitudes make up for about 85% of the total global NBP for the CMIP5 MMM.”

**L 281-294** is informative but needs to be contextualized—currently the authors make a strong case these models are not fit to capture the key feedbacks. Including an objective model evaluation such as an ILAMB broadscore plot would help readers understand “who” these models are. The authors mention that the GFDL model has the best lagged responses, but does it outperform the other models on the benchmarks? This also pertains to the discussion on **L323-333**: The authors mention that CLM5 is the most process rich in the plant hydraulic space, but does it outperform other models in the benchmarks?

We appreciate this suggestion. We understand that the current phrasing of the two paragraphs suggests strong statements that our results do not indicate. Our intention was to discuss potential short comings in process representation in ESMs that may lead to an underestimation of SM-induced changes in the land carbon sink. We rephrased the paragraphs to avoid unjustified strong statements on model performance and clarify that the mentioned study which shows that GFDL capture the lagged effects best compared to observations, serves as an illustrating example. Additionally, we introduced ILAMB broad scores in the second paragraph you referred to in your comment.

We changed the first paragraph to:

“Nevertheless, this does not imply that the overall conclusions of prior studies are overstated, because structural model limitations may lead to an overall underestimation of the severity of SM-induced reductions of NBP in ESMs of both generations. Previous studies highlighted that SM drought can strongly reduce carbon uptake, often followed by declining tree growth rates and increased tree mortality and potentially leading to lasting reductions in carbon uptake capacities of affected ecosystems (Guo et al., 2025; Kannenberg et al., 2024; Kolus et al., 2019). This is often not well captured in ESMs because limitations in the representation of fundamental plant hydraulics can lead to shorter and weaker simulated drought impacts on GPP than observations indicate (Anderegg et al., 2015; Kannenberg et al., 2024; Kolus et al., 2019). For example, some CMIP5 models (including CESM and GFDL, which participated in the GLACE–CMIP5 experiment) do not sufficiently capture the lagged effects of SM on NBP compared to observational data, although GFDL was shown to capture them best of all the CMIP5 models assessed (Anderegg et al., 2015). In addition to the projected strong SM drying, this may further explain the strong negative impact of SM on NBP our study reveals for GFDL, but further research would be needed to confirm this. [...]”

The second paragraph was changed to:

“Among LFMIP models, CESM2 projects the strongest negative SM-induced impact on GPP. Its land surface model, CLM5, explicitly represents plant hydraulics, including leaf water potential, xylem water flow, and vulnerability to hydraulic failure, which are important for regulating the GPP response to dryness stress (Kennedy et al., 2019; Lawrence et al., 2019). In contrast, the land surface models of the other LFMIP models (i.e., CLM4.5, ORCHIDEE v2.0, and JSBACH3.2, see Table S1), rely on empirical water stress functions for stomatal conductance and photosynthesis (Lawrence et al., 2019; MacBean et al., 2020; Reick et al., 2021). According to the International Land Model Benchmarking (ILAMB) broad scores, CESM2 performs well in benchmarks for Global Net Ecosystem Carbon Balance (benchmark related to NBP) compared to other CMIP6 models (Hassler et al., 2026), which might partially reflect the advanced process formulation of CLM5. This suggests that the lack of explicit hydraulic constraints in the land surface models of most LFMIP models may limit their ability to accurately capture the impact of SM on GPP, and that further development of plant hydraulics might improve predictions of GPP responses, especially under conditions of increased moisture limitation. However, despite ongoing efforts, the complex mechanisms regulating ecosystems' response to drought are still not fully understood (Kannenberg et al., 2020), and the observed non-linear and species-specific responses of GPP to dryness stress make it challenging to develop accurate implementations (Green, 2024; Grossiord et al., 2020).”

We further introduced the potential advantages of using ILAMB in the limitation and outlook section of the discussion:

“Using benchmarking frameworks such as ILAMB helps to assess model performance by systematically comparing model output to observational datasets (Collier et al., 2018). Overall,

CMIP6 models show improved skill relative to CMIP5 in representing key components of the carbon and hydrological cycles, indicating that model accuracy has been improved (Hassler et al., 2026). However, water–carbon coupling emerges from tightly interacting processes across soil, vegetation, and atmosphere. As land surface models have grown in structural complexity (Fisher and Koven, 2020), evaluating individual variables in isolation may not reflect how well models capture underlying process interactions. Indeed, even when models perform well for SM and GPP individually according to the ILAMB broad scores, the coupling between SM and GPP may still be misrepresented (Massoud et al., 2025). Expanding benchmarking approaches to explicitly target soil moisture–carbon interactions, dryness stress responses, and regime shifts between energy– and moisture–limited conditions would provide valuable constraints to assess model performance regarding SM–induced effects on NBP.

However, benchmarking simulated SM and its impact on the land carbon sink against past observations can only provide limited insights about satisfactory model performance under climate change in the future, particularly since the effects strongly depend on the forcing scenario and human perturbations of the water and carbon cycles (Gier et al., 2024; Zaitchik et al., 2023). Assessing and improving the representation of water–carbon interactions under emerging climate states therefore remains a key challenge for the next generation of ESMs.”

## Response to Anonymous Referee #2

Dear Referee,

Thank you for taking the time to review this manuscript and for your thoughtful comments. We greatly appreciate your recommendations for improving the paper.

In our responses to the specific comments, we refer hereafter to the line numbers of the marked-up manuscript. In addition, we further corrected a few grammatical errors during revision.

In this study, the authors compared the effects of soil moisture on GPP across CMIP5 and CMIP6 models. They found that SM variability is the main driver of land carbon reductions, and SM-atmosphere coupling is a critical factor for predicting future land carbon uptake. This research is complex and interesting, the manuscript is well-written, and suitable for the journal. The authors have put in a lot of work to complete the manuscript. However, there are some major concerns. Since this paper conducted a similar analysis to Green et al. (2019) and the main findings are similar, it is unclear what the innovative part of this study is. Secondly, the analysis is comprehensive and thorough, but missing uncertainty analysis, especially for the spatial maps (only shows the mean, without a uncertainties map). Thirdly, the method section needs significant improvement, especially regarding how the selected model simulates soil moisture, GPP, and how temperature affects GPP. Moreover, the authors did not explain whether the soil biogeochemical component of the model is included, as it is also an important part of carbon cycling. Did these models have the soil biogeochemical part? How will soil moisture affect microbial activity and Nitrogen availability? There is also a lack of a mechanism explanation for findings, please try to explain the mechanisms. Lastly, there are so many acronyms in the manuscript, which makes it very difficult to follow. Suggest making a list/table of the different acronyms. As a result, I would like to suggest a major revision.

Thank you for the overall positive evaluation of our work. Your suggestions are very helpful to include additional uncertainty information, clarify the process representation in the considered LFMIP/CMIP6 models, and improve overall readability and organisation. Detailed responses are provided with the specific comments below.

On a general level, we would like to address the concern regarding the innovative part of this study compared with Green et al. (2019). In the first part of our manuscript (section 3.1), we analyse the results for the latest version of the SM experiment, LFMIP (based on CMIP6 models), and compare it with those for the previous version, GLACE-CMIP5 (based on CMIP5 models). To our knowledge, this is the first analysis of SM-induced impacts (trend and variability) on the land carbon sink using LFMIP. Furthermore, the comparison between CMIP5 and CMIP6 model generations has not been done before. The results are surprising as they demonstrate great differences in the magnitude of SM-induced impacts on the land carbon sink. This led us to the second part (section 3.2) where we investigate the origins of SM-induced changes in land carbon uptake (GPP), and whether intermodel differences are driven by differences in the SM evolution or the sensitivity of GPP to SM. This was not included in the study of Green et al.

(2019). To clarify these contributions, we will formulate research question at the end of the introduction:

1. Is the SM-induced impact on the land carbon sink (NBP) according to LFMIP (CMIP6) different from that according to GLACE-CMIP5 (section 3.1)?
2. Within LFMIP (CMIP6), what are the origins of (i) SM-induced changes in land carbon uptake (GPP) (section 3.2.1) and (ii) related intermodel differences (section 3.2.2)?

Make a clear table showing the relationship between LFMIP and GLACE-CMIP5, as many acronyms for models or projects are making the manuscript very hard to read.

We appreciate this suggestion. A table providing an overview over the experiments and included models is a very good idea. Such a table will also give an overview of most of the acronyms used throughout the paper and we believe this will make a dedicate acronym list obsolete.

A table was added (Table1) in the method section to give a short overview on the experiments LFMIP and GLACE-CMIP5. It includes information on experiment names (GLACE-CMIP5/LFMIP), CMIP generation (CMIP5/CMIP6), and available models (listing the four models available for GLACE-CMIP5 and LFMIP (CMIP6), respectively). The table is mentioned in the text in line 109.

Line 30: The concept of NBP is a little unclear, please elaborate more of this concept and what is the difference between NBP and NEP?

The difference between NBP and NEP is that NBP also includes carbon loss due to disturbances (i.e.,  $NBP = NEP - \text{disturbances}$ , resulting in  $NBP = GPP - \text{ecosystem respiration} - \text{disturbances}$ ). Consequently, NBP expands the concept of NEP to larger spatial and temporal scales, explicitly including disturbances. Keenan and Williams (2018), which we refer to where NBP is introduced, provide detailed definitions for carbon sink related terms such as NEP and NBP. Introducing NEP in our paper would perhaps exceed the scope, as we analyse either NBP or GPP (but never NEP).

Nevertheless, we understand that the current phrasing of the paragraph is confusing, and we edited the paragraph to make the concept of NBP clearer.

The sentence was changed to:

“The terrestrial biosphere plays a crucial role in Earth's climate system by acting as a net carbon sink. The land carbon sink is determined by the balance of carbon uptake through photosynthesis (Gross Primary Production, GPP) and losses via ecosystem respiration and disturbances such as wildfires and land–use change (Keenan and Williams, 2018). On large spatial and temporal scales, this net carbon balance is referred to as Net Biome Production (NBP), where positive values indicate a net flux of carbon from atmosphere to land (sink) and negative values indicate a net flux of carbon from land to atmosphere (source).”

Line 35: temperature-induced increases in respiration, plants or soil microbials?

In this line, “temperature-induced increases in respiration” refers to both, plant respiration and soil microbial respiration. Therefore, the sentence was rephrased to: “temperature-induced increases in respiration by plants and soil microbials, [...]”.

A brief explanation on how SM affects respiration (plant and soil microbial respiration) was added in the introduction:

“ [...]. In many ecosystems globally, reduced SM has a more pronounced negative impact on NBP through GPP than respiration (Gu et al., 2017; Li et al., 2023; Von Buttlar et al., 2018). Nevertheless, ecosystem respiration is also sensitive to water scarcity. Low SM can suppress soil respiration by reducing the availability of soluble carbon substrates to soil microbes, while elevated temperatures can lead to higher soil respiration (Gentine et al., 2019).”

Line 50: Does this research include the below-ground soil-biogeochimistry or only the plant-vegetation part?

All models participating in LFMIP (CMIP6) and GLACE-CMIP5 are fully coupled Earth System Models (ESMs) which all include both, below-ground soil-biogeochimistry and above-ground vegetation. A paragraph was added in the method section on ‘Model experiments and data’ to briefly describe how these mechanisms are represented in the models (line 126 to 130):

“ESMs are physically based, prognostic models that include a land component with an interactive carbon cycle described by numerical representations of real-world biogeochimical processes. GPP is simulated as leaf level photosynthesis and constrained by temperature, radiation, atmospheric CO<sub>2</sub> concentration, water and nutrient availability. ESMs also represent soil biogeochimistry, including litter and soil organic matter decomposition that drive microbial respiration. NBP is provided as the net land-atmosphere carbon exchange (i.e.,  $NBP = GPP - \text{ecosystem respiration} - \text{disturbances}$ ).”

Line 57, this part is unclear, by further reducing SM, through what feedback loop? Higher transpiration or higher evaporation? Higher VPD may also cause a close of stoma to limit water loss, so it still decreases transpiration which may not reduce SM.

We apologize for the confusion; this paragraph is poorly phrased. We rephrased it to clarify the concept of indirect SM effects via temperature, VPD, and radiation on land carbon uptake.

The paragraph was rephrased to:

“SM limitation impacts both major carbon sink components, GPP and ecosystem respiration. Decreased SM can cause water stress in vegetation due to direct physiological effects of water limitation on GPP and indirect effects via land–atmosphere feedbacks (Green et al., 2019; Liu et al., 2025; Zhou et al., 2019). Direct SM effects arise from photosynthetic activity being constrained by soil water availability. Under dry conditions, plants partially close their stomata to limit water losses through transpiration, which in turn reduces CO<sub>2</sub> uptake and thus limits GPP (Sippel et al., 2018). Indirect SM effects arise from interactions between the land surface and the atmosphere. Low SM suppresses latent heat flux and evapotranspiration, leading to higher surface temperatures and vapour pressure deficit (VPD). Low SM suppresses the latent heat flux through evapotranspiration, causing a shift towards sensible heat flux that increases near-surface air temperature and vapour pressure deficit (VPD). Elevated temperatures and increased VPD enhance atmospheric evaporative demand and promote further stomatal closure, thereby reinforcing the direct limitation of low SM on GPP (Berg et al., 2016; Seneviratne et al., 2010). Low SM can also increase incoming radiation by reducing cloud formation and atmospheric humidity, which enhances photosynthetic activity (Seneviratne et al., 2010, Zscheischler et al., 2015). In many ecosystems globally, reduced SM has a more pronounced negative impact on NBP through GPP than respiration (Gu et al., 2017; Li et al., 2023; Von Buttlar et al., 2018). Nevertheless, ecosystem respiration is also sensitive to water scarcity. Low SM can suppress soil respiration by reducing the availability of soluble carbon substrates to soil microbes, while elevated temperatures can lead to higher soil respiration (Gentine et al., 2016).”

Suggest moving table S1 to the main text and making the ESM the first row instead of using the modelling group as the first row. What is the land model resolution? How many soil layers does each model have?

Table S1 can be moved to the main text, however, we now added Table 1 to the main text, which gives an overview on the experiments LFMIP and GLACE-CMIP5. This is why we decided to keep Table S1 in the supplement.

For all models in Table S1, the land model resolution is the same as the atmospheric resolution. The row was renamed to ‘grid resolution’ to make this clear.

As we use standardized SM values (i.e., z-score) in our analysis and the focus lies on the effect of SM changes on the land carbon sink and not on the SM changes themselves or soil processes, we believe that including the number of soil layers is not of great importance for the analysis.

Line 100, do not understand this paragraph, what is the purpose of selecting one ensemble member? Did the authors already select four models in the previous description?

For the experiments GLACE-CMIP5 and LFMIP (CMIP6), we do not select models. Only four modeling groups participated with their respective model and ran one simulation (i.e., member) per experiment. Given the small number of available models, we use all four available models and evaluate their characteristics (e.g., SM, NBP) with respect to the CMIP ensemble consisting of 9-10 models. For this comparison we use the control run, which is identical to future projections simulated by all CMIP models (in contrast to the experiments with SM prescription).

Line 120 refers to the full ensemble of CMIP models, where multiple members (i.e., model runs) are available per model. We check if the available four models for the LFMIP/GLACE experiments (with only one member available) are representative for the larger CMIP5/CMIP6 ensembles using one member per model for comparability.

We improved the text to clarify this and changed the paragraph to:

“To assess if the available models of LFMIP and GLACE–CMIP5 are representative of their respective CMIP generation, larger sets of CMIP6 (Eyring et al., 2016) and CMIP5 (Taylor et al., 2012) models are used. Ten ESMs of CMIP5 (i.e., BNU–ESM, CESM1–BGC, CanESM2, CanESM2, GFDL–ESM2G, HadGEM2–CC, HadGEM2–CC, IPSL–CM5A–LR, MPI–ESM–LR, and NorESM1–M) and nine ESMs of CMIP6 (i.e., ACCESS–ESM1–5, CESM2, CMCC–ESM2, EC–Earth3–Veg–LR, GFDL–ESM4, IPSL–CM6A–LR, MPI–ESM1–2–LR, MRI–ESM2–0, NorESM2–LM) were used for this comparison, selecting one ensemble member from each modelling group.”

Line 105, how is the NBP calculated?

The models output NBP as a standard diagnostic (i.e.,  $NBP = GPP - \text{ecosystem respiration} - \text{disturbances}$ ). As mentioned above, in the method section we added a brief explanation on how these mechanisms are represented in the models.

Figure 1, there is a mismatch of caption descriptions and figure panels, please check and revise. The label of y-axis is a little confusing.

Thank you for pointing this out. The caption of panel (a) and (b) were indeed swapped, this was corrected. To avoid confusion on the y-axis label, we will make clear in the description (caption and title) of panel c) that it displays a 30-year mean. Consequently, all data is shown in the same unit ( $\text{Pg C yr}^{-1}$ ).

Line 181, “reduces the global land carbon sink to half”, this sentence is confusing and difficult to understand.

We understand that this is confusing, and we added a brief explanation in brackets to clarify the meaning of this: “The total impact of SM on NBP at the end of the century reduces the global land carbon sink to half (i.e., the land carbon sink would be twice as strong without the negative impact of SM).”

Figure 1, why do you choose 20 20-year centred rolling mean and your analysis period is 29 years

For the time series in panel a) and b) of Figure 1 we chose a 20-year rolling mean to remove the interannual variability and focus on the long-term trend without losing the features of the time series (which would happen when smoothing over longer periods). For the analysis of the impact of SM on NBP by the end of the century (panel c) we chose a 30-year period for averaging. A 30-year period is usually used to define “climate normals” because it is long enough to capture natural variability without being dominated by short-term weather fluctuations, and short enough for updating trends. In summary, 20 years are chosen for calculating trends and 30 years for calculating end of the century averages.

Figure S4 is never mentioned or explained in the main text.

Thank you for pointing this out. Figure S4 is now mentioned in the results section in the first paragraph (line 198).

Lines 205 – 210, the spatial analysis shows that the large SM differences between CMIP5 and LFMIP for the IPSL and GFDL models. Can you explain what is causing the differences. Moreover, if the ensemble of many models means kind of similar, this indicates the IPSL and GFDL may not be typical representative models, so why do you select these two models?

Thank you for this question. As mentioned under an earlier comment, we do not select models but use the four available models for GLACE-CMIP5 and LFMIP, respectively, and evaluate their characteristics (e.g., SM, NBP) with respect to the CMIP ensemble consisting of 9-10 models. This was now clarified in the methods.

When comparing SM of IPSL and GFDL against a larger set of CMIP5 models we find that they are not representative of the CMIP5 multi-model mean. However, this does not mean that these models are not plausible and should be removed. The purpose of the comparison is to put the available models into context and provide an overview on where they lie within the ensemble (e.g., are they among drying or wetting models, what do their spatial patterns of drying/wetting look like). This is important information to consider when interpreting the results. Furthermore, by comparing SM within and across model generations, we acknowledge that models and

generations differ. The newly added Table 1 now clarifies which models are available for the experiments, GLACE-CMIP5 and LFMIP, and to what CMIP generation they belong. Together with the reorganized and more structured method section, this will hopefully clarify the confusions concerning the models used in this analysis.

Why the models project different SM evolution is outside the scope of our work as our focus lies on SM-induced impacts on NBP and not on SM itself. However, other papers investigate that in more detail, e.g., Berg et al. (2017) – “Divergent surface and total soil moisture projections under global warming” and Liu et al. (2024) – “Evaluation of 22 CMIP6 model-derived global soil moisture products of different shared socioeconomic pathways”. We refer the reader to these references in the discussion section on limitations and outlook.

Figure 2 and 3 may need to perform some uncertainty analysis of the spatial patterns.

Thank you for this comment. This suggestion was implemented for Figure 2 and 3. As our analysis focuses on two 30-year time periods, namely the baseline (1981-2010) and the future period (2070-2099), we indicated significant changes in the future period relative to the baseline by adding stippling in the model-specific plots. For model mean plots, we added hatching for sign disagreement. We extended the captions to include this information. In addition, we applied the same approach to related Figures in the supplement (Supplementary Figure S5 to S7).

Lines 212, be specific, how many available models? Please also describe this more clear in the method section.

We apologize for the lack of clarity. We added “of the **four** available models” for the GLACE-CMIP5 and LFMIP and rephrased the paragraph such that it becomes clearer:

” We assess the representativeness of the four available models in GLACE-CMIP5 and LFMIP for the respective CMIP generation by comparing them with multi-model means (MMMs) of CMIP5 and CMIP6, respectively. The comparison was performed for NBP and SM for each latitude zone [...]”

Figure 5, please include the meaning of  $GPP_R$  in the caption too. Please also explain in the text why the indirect effect R from SM leads to an increase in GPP. Can the selected model capture the optimum temperature effect for plant GPP.

The caption was edited such that the meaning of each term, including  $GPP_R$ , is clear. Furthermore, we explained in the method section how indirect SM-effects via radiation can increase GPP:

“Low SM can also increase incoming radiation by reducing cloud formation and atmospheric humidity, which enhances photosynthetic activity (Seneviratne et al., 2010, Zscheischler et al., 2015).”

The models capture the effect of temperature on GPP, since photosynthesis is implemented as a function of radiation, temperature, water (and nutrient) availability and CO<sub>2</sub>. As suggested in another comment, we added a brief explanation on how the variables of interest are simulated when introducing the models in the methods section. However, this study does not intend to conduct a technical comparison of process implementation, and a more detailed comparison would exceed the scope of this analysis. In table S1 we list references for each model, which provide further information on how the models are built, with detailed explanations of how specific processes are implemented conceptually and mathematically in each model.

Lines 315-320, please be specific, “some land surface models” which models? Maybe in the method section describe a little how the GPP and soil moisture are modelled in the four selected models.

We agree that this formulation sounds vague and needs to be rephrased. It does not refer to specific models used in this analysis, but to state-of-the-art land surface models in general (as shown by Franks et al., 2018), which are also used in LFMIP models. The aim is to point out that the process formulation of stomatal conductance in state-of-the-art ESMs is often linked to VPD. This could be one explanation why indirect effects are found to be a major driver of SM-induced changes of GPP in our analysis. We edited this part for clarity and rephrased it to: “In state-of-the-art land surface models, the stomatal conductance formulation is often directly linked to VPD (Franks et al., 2018). [...]”.

As suggested, we also added a brief explanation on how the variables of interest are simulated when introducing the models in the method section (see response above).

Line 369: How to quantify the SM variability quantitatively?

The existing GLACE-CMIP5/LFMIP experimental setup allows isolating the model-specific effect of SM trend and variability on the land carbon sink (from any other future changes in the land carbon sink). However, it does not allow separating the effects of differences in SM evolution and sensitivity to SM across models. Therefore, in this paragraph we suggest additional experiments with identical SM evolution, e.g., using the CMIP MMM of soil moisture (corresponding to SM in the CTL experiment). Based on that, an identical SM trend (similar to rmLC) and an identical prescribed mean seasonal cycle of SM (similar to pdLC) could be implemented. Such experiments enable to isolating the effect of identical SM trend and variability on the land carbon sink which would help to assess intermodel differences in SM-

induced changes of NBP in further detail. We understand that the current phrasing is too short, and we therefore elaborated further to better explain this idea.

The paragraph was rephrased to:

“Conducting GLACE-like experiments with a larger set of models would enhance the robustness and validity of the results and provide further insights into the effects of SM on the land carbon sink and associated uncertainties under future climate change. We therefore recommend continuing GLACE-like experiments in future CMIP generations with more models to generate a larger ensemble. Furthermore, it has been shown that SM evolutions differ substantially across ESMs, due to differences in model process implementation (Berg et al., 2017, Liu et al. 2024b), which complicates intermodel comparison. We therefore propose to extend the experimental setup by an experiment with identically prescribed SM evolution (e.g., prescribed trend and variability based on the CMIP MMM). This would enable isolating the effects of identical SM trend and variability on NBP (while the existing GLACE-like experimental setups allow isolating only the model-specific effect of SM trend and variability), by explicitly disentangling the contributions of differences in model sensitivity of GPP to SM from intermodel deviations in the simulated SM itself. Since our results from the factorial ANOVA indicate that differences in the sensitivity of GPP to SM are the main origin of intermodel uncertainty, this approach would allow for a more direct comparison of the representation of water-carbon coupling across ESMs.”