Response to Reviewer 2

Lee and Hohenegger investigate the hypothesis that differences in precipiation patterns and frequency are the dominant cause for carbon cycle differences between coarse and high resolution climate models. They analyse outputs of one land surface model driven by output from two different climate models, which also operate of different spatial resolutions. While the hypothesis is interesting, the manuscript suffers from a number of major and minor flaws. I am uncertain whether these can be addressed in a satisfactory manner by revisions.

We thank the reviewer for his/her thorough evaluation of our manuscript and valuable comments, which allowed to significantly strengthen our conclusions. To be able to properly answer the reviewer concerns, we conducted extensive new analysis and performed several new simulations. The goal was to rule out the role of confounding factors in the GPP-precipitation relationship and to better isolate which aspects of our results are robust and may generalize and which not, as asked by the reviewer. Please find our detailed, point-by-point responses to all major and minor comments below.

1. The authors acknowledge that the climate field used as input differ in nearly all aspects of climate (temperature, radiation, humidity, rain), however, their method to disentangle these co-variate effects) is at best cursory. Figure 2, for instance, analyses the dependency of simulated gross primary productivity to precipitation intensity, without controlling for covarying factors such as average and seasonality of temperature, air humidity etc. As the authors do not provide any analysis of the seasonality of precipitation intensity, it is impossible to evaluate to what extend seasonal variations in temperature (and possibly phenology) confound this finding.

To address the reviewer's concern, we conducted two main supplementary analyses. First, to isolate the drivers of the GPP differences, we conducted three additional offline sensitivity experiments using LR configuration but selectively substituting specific high-resolution (SRM) atmospheric forcing: (a) precipitation, shortwave radiation, and air temperature; (b) precipitation and shortwave radiation only; and (c) SW only.

As shown in the Figure R2.1 below, replacing radiation only is sufficient to remove the decline in GPP with intensifying precipitation in LR. This isolates shortwave as the primary, proximate driver. Importantly, this strongly supports our proposed mechanism: the suppression of weaker and longer-lasting event in the SRM forcing (for the same amount of precipitation) results in less persistent clouds, allowing significantly more downward shortwave radiation to reach the canopy during the growing season. Thus, the difference in precipitation characteristics fundamentally alters the radiation regime, which in turn prevents the GPP decrease with more intense precipitation events.

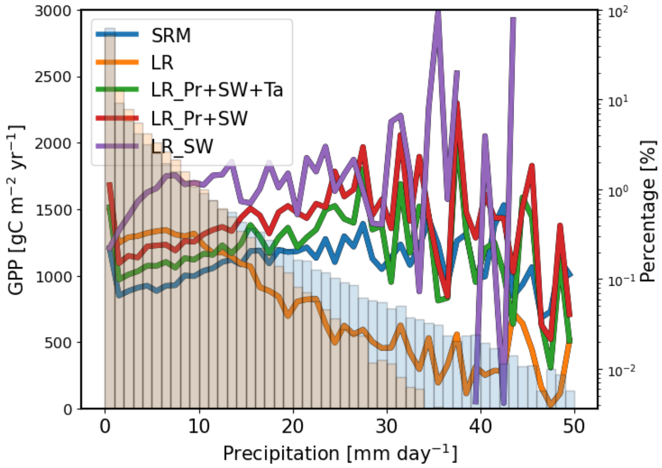

Fig. R2.1*: Daily mean GPP as a function of daily mean precipitation for LR (orange line), SRM (blue line), and three LR sensitivity experiments in which selected atmospheric forcing variables are replaced by their SRM counterparts: precipitation (Pr), downward shortwave radiation (SW), and air temperature (Ta) in LR_Pr+SW+Ta (green); Pr and SW in LR_Pr+SW (red); and SW only in LR_SW (violet). The bars show histograms of precipitation for LR and SRM. Note that in the revised version, we changed our baseline LR simulation (see response to comment 2 of the reviewer). We conducted these sensitivity experiments on both baseline LR, it does not matter, here we show the results based on the new LR baseline simulation.*

Second, to address the reviewer's concern regarding seasonal confounding factors, we repeated the analysis of the GPP–precipitation intensity relationship for the four seasons (DJF, MAM, JJA, and SON) individually. As shown in Fig. R2.2 below, the difference in GPP between the two models for high intensity precipitation events is only apparent in summer (JJA). This is consistent with our hypothesis, as if the decline in the GPP-precipitation relationship in LR results from the treatment of convection, it should be visible in the summer season. It is also the only season where more than half of the precipitation is produced by the convection scheme. The contributions are: 12% in DJF, 31% in MAM, 56% in JJA and 23% in SON.

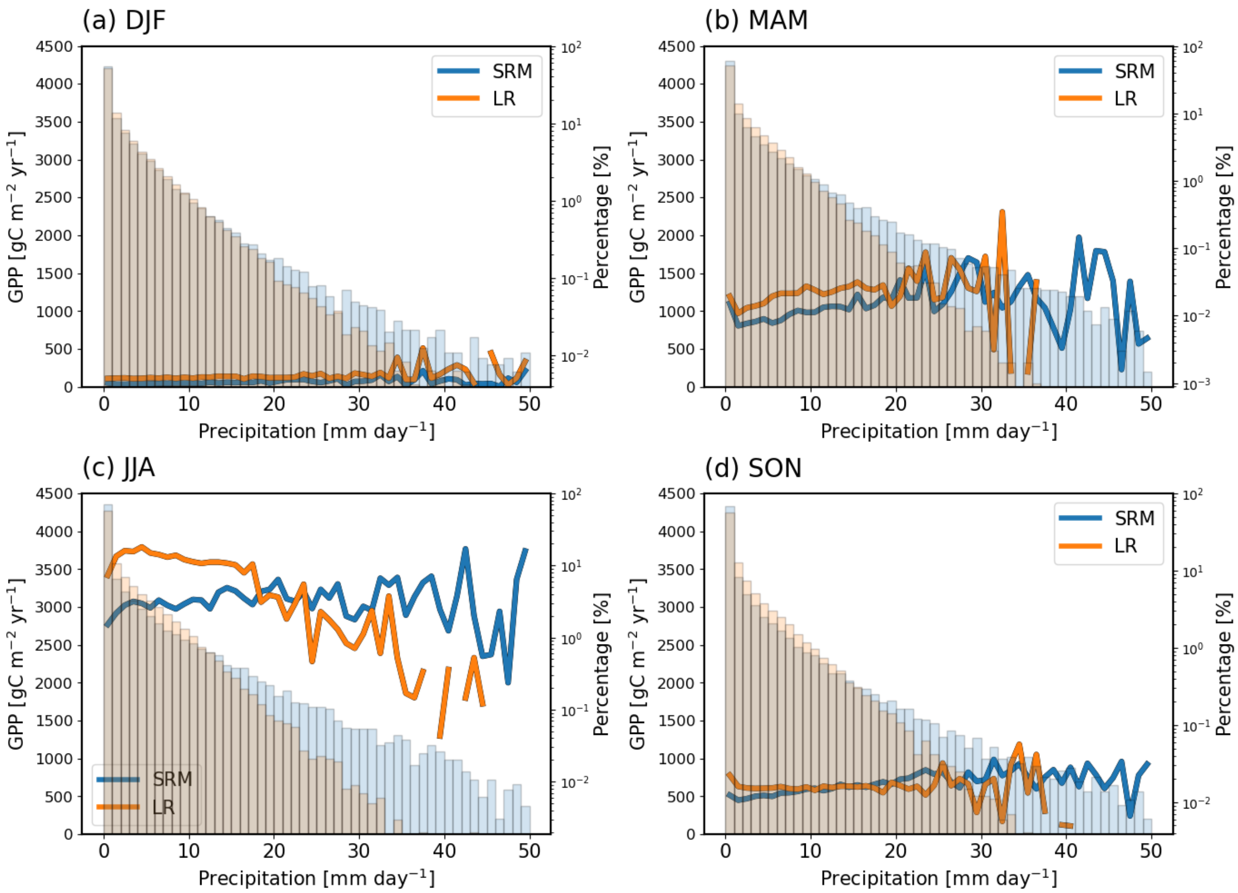

Fig. R2.2: *Daily mean GPP as a function of daily mean precipitation for LR (orange line) and SRM (blue line), together with their histogram of precipitation (bar), for (a) December-February (DJF), (b) March-May (MAM), (c) June-August (JJA), and (d) September-November (SON). Note that in the revised version, we changed our baseline LR simulation (see response to comment 2 of the reviewer). We conducted these sensitivity experiments on both baseline LR, it does not matter, here we show the results based on the new LR baseline simulation.*

Moreover, in response to one further comment of the reviewer (see for more detail comment 2 of the reviewer), we reran the LR configuration for a different climate, using six different ensemble members. The decline of GPP by intensifying precipitation remains in the LR (see Fig. R3a). This is so, because the relationship between SW radiation and precipitation remains identical as in our baseline simulation (see Fig. R3b). Likewise, also in that simulation, the strong decrease in GPP with increasing precipitation amounts is visible in MMA and JJA, in agreement with those being the more convective seasons.

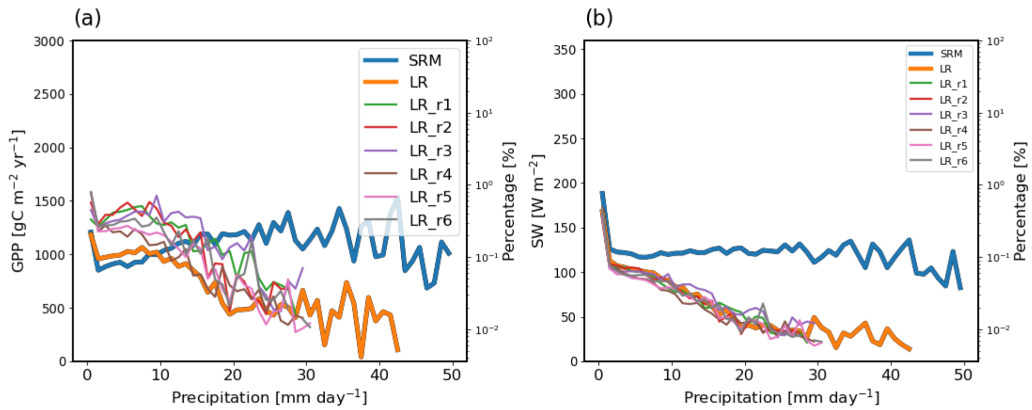

***Fig. R2.3**: (a) Daily mean GPP as a function of daily mean precipitation for LR (orange line), SRM (blue line), and new LR baseline simulations (LR_rX). The bars show histograms of precipitation for LR and SRM. (b) Same for daily shortwave radiation as a function of daily mean precipitation.*

All in all, we believe that this supplementary analysis strengthens and strongly supports our initial results that GPP declines too rapidly in LR under high precipitation events due to the treatment of convection and the strong decrease in radiation. We added the discussion of the results of the new sensitivity experiments in section 3.2.1 on lines 217-225 and added Fig. R2.1 as supplementary Fig. S1. We also added the results of the seasonality analysis in section 3.2.1 on lines and added Fig. R2.2 as supplementary Fig. S2. Finally, we now use the new LR baseline simulation (for reasons explained in the next comment) and correspondingly updated Tables, Figures and text.

2. The authors interpret small (a few percent) regional differences in average carbon fluxes in terms of the hypothesised effect of rainfall intensity, but fail to ascertain that the difference do not simply stem from difference in the spatial extend of land mass resulting from the obviously radically different land-sea mask. No evaluation is presented that rules out that difference presented are simply an expression of slight shifts in the contribution of different plant functional types resulting from the different resolutions. No assessment is presented in terms of the different average climatologies of the simulations. It is for instance very conceivable that the strong different in East-European gross primary production (GPP) between the two simulation is an expression of different seasonal or annual mean precipitation and thus moisture stress, rather than of the frequency of rainfall per-se. However, an evaluation of these climate characteristics and spatial patterns is not given.

The reviewer raises concerns regarding potential confounding factors such as 1) land-sea mask differences, 2) PFT differences, and 3) spatial climatology. We address each of these points below to demonstrate that the hypothesized effect of rainfall intensity remains robust.

1): In order to completely rule out artifacts stemming from the different spatial extents of the land-sea masks between SRM and LR, our entire analysis in the original manuscript was already strictly confined to grid boxes that are defined as 100% land in both simulations. We agree that we should have mentioned this, which we failed to do in the original manuscript, and we updated the method section accordingly to clarify this aspect, see lines 130-131.

2): As explained in the method, the probability density functions of both simulations are derived from the same high-resolution (300 m) land cover data set, which is then coarsened to the respective grid spacing. This coarsening can indeed lead to small differences between the simulations. For instance, SRM has larger C3 grassland and shrub at the expense of C3 cropland and deciduous tree, with a fraction difference of ~1% over the entire domain. To ascertain that our GPP results are not simply an artifact of such PFT differences, we conducted a new analysis where we look at the GPP-precipitation relationship for a given PFT. We do this analysis for C3 cropland, C3 grassland and extra-tropical evergreen, which are the main dominant PFTs in our domain. We require a coverage larger than 80% of that dominant PFT in a grid box, as no grid box exists with a dominant PFT occupying more than 80%. The result is shown in Fig. R2.4 below. The decline of GPP with intensifying precipitation is visible for all pfts. The SRM and LR curves are most similar for grassland. As grassland occupies a slightly larger fraction of the domain in LR, if anything, this should have made the domain-mean GPP-precipitation curves of the two models more similar. Hence differences in in the fractions of plant functional types between SRM and LR can also not explain the distinct behavior of GPP with intensifying precipitation between the two models. This is consistent with the finding that using the SW forcing of the SRM for the LR is sufficient to “correct” the precipitation-GPP relationship in the LR (see our response to reviewer comment 1). We added the results of this analysis in section 3.2.1, lines 232-237, and Fig. R2.4 as Fig. S3 in the supplementary information.

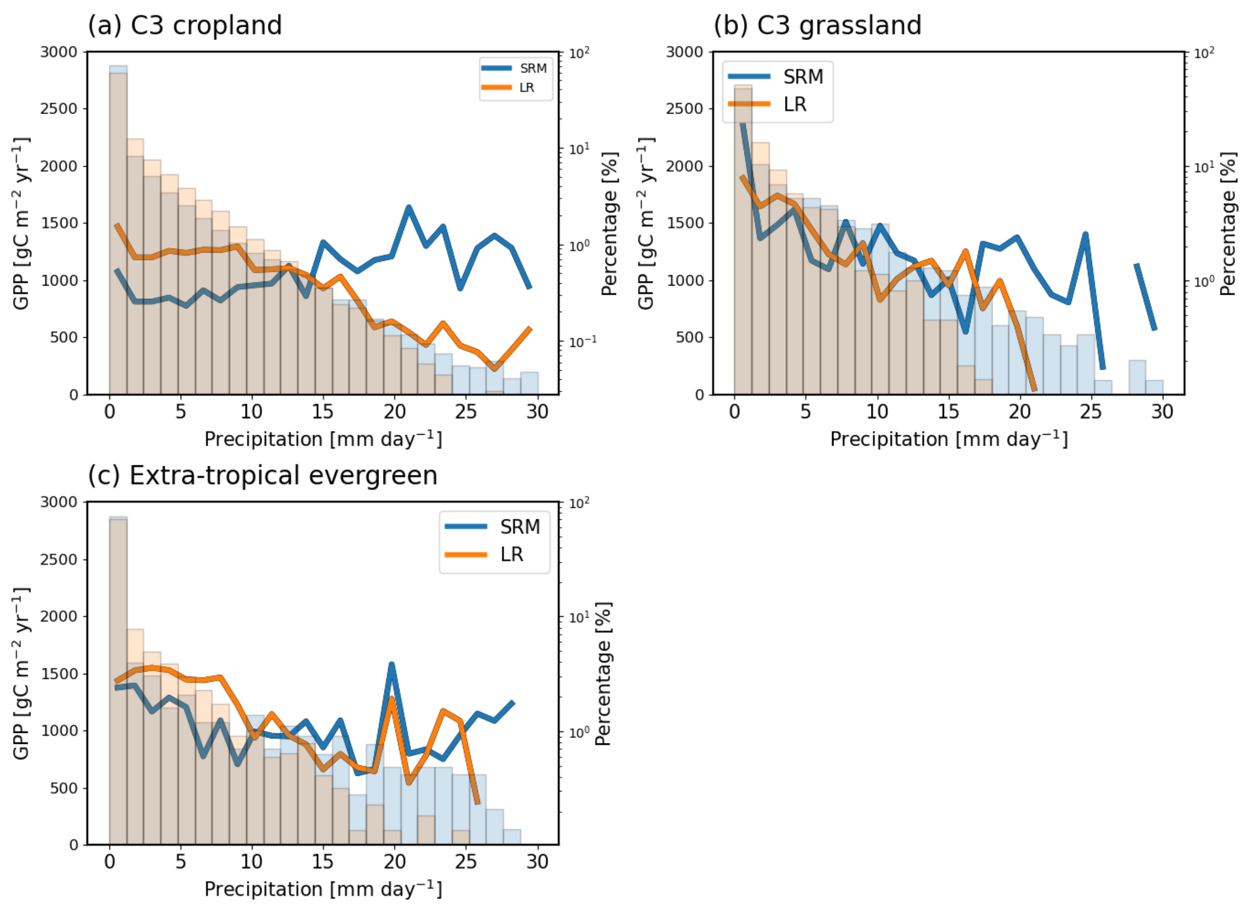

***Fig. R2.4**: Daily mean GPP as a function of daily mean precipitation for LR (orange line) and SRM (blue line), together with their histogram of precipitation (bar), for grid cells dominated by (a) C3 cropland, (b) C3 grassland, and (c) Extra-tropical evergreen. Note that in the revised version, we changed our baseline LR simulation. We conducted these sensitivity experiments on both baseline LR, it does not matter, here we show the results based on the new LR baseline simulation.*

3): The reviewer suggests that the strong differences in East-European GPP between SRM and LR might be an expression of annual mean precipitation and moisture stress difference between the two models, and notes that such an evaluation was missing. We totally agree with the reviewer's interpretation, as this was also our interpretation, and the corresponding analysis and interpretation was already reported in the original manuscript, see lines 8-10 in the abstract, lines 202-220 in the main text, corresponding Fig. 3, and lines 311-320 in the conclusions.

In the original manuscript, we advocated that the difference in the GPP-precipitation relationship by high rain intensity events is a fundamental difference between LR and SRM models (explicit versus parameterized convection), whereas the resulting mean change is modulated by the role of the mean climatology in determining GPP. For instance, at the end of the conclusions, we had written “Hence, although explicitly resolving convection favors higher gross primary productivity through shorter-lived, intense daily precipitation events, the overall climatology modulates this effect”. Given this influence, which we especially saw over eastern Europe with an associated decrease in GPP, we realized that it was an unfortunate decision in the submitted version that we compared two different basic climates, using multi-year “2020” climate from the SRM with multi-year piControl climate from the LR. At the time of the study, the multi-year “2020” climate from the SRM was the only SRM simulation available, and having multi years of the same climate was seen as useful to sample internal variability. As we didn’t have an exactly corresponding set-up at LR, we used piControl (which has many years of the same climate). For this revision, we simulated again LR taking the atmospheric forcing from the MPI-ESM-LR simulation of the year 2021 of the ssp2-4.5 scenario, as that year had about the same CO_2 concentration in the atmosphere than SRM (414.4 versus 414.39), hence allowing us to compare in principle similar climates. We then used six different members to sample internal variability. How does this change affect our previous findings?

In a nutshell, the key previous findings of our study were that (i) the pdf of rain affects GPP and explicitly resolving convection prevents a strong decrease of GPP by higher intensity precipitation events (for a given precipitation amount); (ii) in our case, this led to an increase in mean GPP but the resulting mean change depended on the simulated climate, especially a tendency for drier conditions over eastern Europe which counteract the increase in GPP due to shorter-lived intense precipitation events; this also explained the E-W gradient in GPP difference we obtained; (iii) consistent with the GPP changes, respiration changed with a larger Ra and a smaller Rh, the latter due to drier conditions.

The new conducted simulation confirmed (i). It also confirmed (ii), in the sense that the mean GPP difference depends on the simulated climate, but in this case, in the mean, GPP decreases as the drying in the SRM atmospheric forcing compared to the LR forcing is more widespread and even more pronounced over eastern Europe; the E-W gradient in GPP difference remains. It is actually remarkable that the drying over eastern Europe is still present, this seems to be a fundamental difference in the atmospheric forcing between SRM and LR. Finally, concerning (iii), as Ra still follows the GPP changes in this new simulation, it has to be (and is) smaller in the SRM. Rh remains smaller (see our response to reviewer comment 3). All in all, and incidentally, we believe this allows us to now convincingly isolate what is robust in the difference, what is not, and what we should expect in general for GPP between km and coarse-resolution atmospheric forcing, which were also the main concerns of the reviewer. In the revised version, we used this new LR baseline and correspondingly updated text, Figures and Tables.

3. The authors imply that the decline in heterotrophic respiration (Rh) between coarse and high resolution is a function of rain fall intensity. However, they fail to evaluate the more likely reason that the stronger increase in autotrophic respiration (Ra) compared to GPP results from a shift in carbon allocation (the split between respiring and non-respring biomass), which potentially results from small shifts in the relative contribution of herbaceous and wood PFTs to continental productivity. Given that the authors spun the model into equilibrium for both configuration, the stronger increase in Ra over GPP requires a decline in Rh for the carbon budget to be balanced.

We agree (as also mentioned in the manuscript) that the total carbon budget must be balanced. Compared to our new baseline LR, SRM exhibits smaller GPP, smaller autotrophic respiration and smaller heterotrophic respiration.

In response to the reviewer comment, we investigated whether changes in carbon pools could explain the decline in Rh. In JSBACH, the soil carbon pools are non-woody litter, woody litter, each subdivided into ethanol, acid, water and nonsoluble with their own decomposition rates. The key difference is that the decomposition rates in the woody litter are multiplied by a factor to account for pore size, about 0.53. Moreover, there is a humus pool. Looking at the different pool sizes, we find that the humus pool is always larger in SRM than LR, in mean by 3.8 Pg C. In contrast, the litter pool is always smaller in SRM than LR, by 3.3 Pg C. Even though as net effect the total soil carbon pool is larger in SRM than in LR, since the decomposition rate of the humus pool is at least a factor 10 smaller than the ones of the litter pool, the different pool sizes between SRM and LR could still lead to an overall reduction of Rh, depending on how the woody and non-woody litter changes. We computed the subdivision of the litter pool in its two components only once per month and only for one year/one member of the SRM and LR simulation sets as this computation requires writing out all the carbon pools on all plant functional types for every grid point and adds 220 variables to the output. As the pool sizes do not exhibit large temporal variations, this seems sufficient. The non-woody litter is again slightly larger in SRM, with a mean value of 29.4 Pg C versus 29.2 Pg C in LR, whereas the woody litter is smaller in SRM, with a value of 10.6 Pg C versus 13.3 Pg C in LR. The smaller woody litter pool in SRM explains the smaller litter pool in SRM, but the decomposition rates of the woody litter pool are only half the values of the ones of the non-woody litter pool and the pool size is also only about half the non-woody one. Factoring in the pool specific decomposition rates, but neglecting any potential change in the climate parameter by setting it to 1, it turns out that the overall respiration should be larger in SRM due to the larger humus and larger non-woody litter pools. Hence, we conclude that the smaller respiration in SRM is primarily due to climate effects. We added this new analysis at the end of section 3.2.3, see last paragraph.

4. It is questionable that the results obtained by the authors are applicable to any other Earth system model. Although the JSBACH model accounts for the diffuse radiation in the canopy radiation scheme, this information is not used to account for the effect of diffuse radiation on gross primary production. Thus, unlike other land surface models, JSBACH does not account for the change in light use efficiency with cloud cover and hence likely overestimates the GPP sensitivity to shortwave downward radiation.

We agree that JSBACH does not partition the incoming shortwave radiation at the top of the canopy into its diffuse and direct component to account for the distinct effect of diffuse and direct radiation on GPP, even under the same amount of total radiation. However, we want to remark that:

1) The effect of light quality on photosynthesis remains absent, simplified, or insufficiently evaluated in many land-surface models. Zhang et al. (2020), for example, stated that the impact of accurate consideration of diffuse light fraction is not represented in most land-surface models and identified only a limited number of models or specially modified model versions that explicitly distinguish the photosynthetic effects of direct and diffuse radiation. Their diffuse-aware ORCHIDEE version was itself developed by adding this process to the standard ORCHIDEE model. The omission of diffuse-light fertilization is therefore an important but relatively widespread structural limitation, rather than a peculiarity of JSBACH. Moreover, parameterizing this effect is difficult due to the limited number of sites measuring all the required variables.
2) It is well accepted that coarse-resolution models represent the wrong distribution of precipitation rates and that they underestimate short-lived events at the favor for longer-lasting events. It seems hard to believe that coarse-resolution models will have the correct representation of diffuse to direct radiation given that they simulate the wrong distribution of precipitation event durations. Chakraborty and Lee (2021) noted large discrepancies in the fraction of diffuse to direction radiation in reanalysis and that such an evaluation is missing for CMIP6. Hence, it seems that this speaks even more for rising attention to the potential effects of misrepresenting convective events duration on GPP at coarse resolution and the need for further studies. It would be unfortunate to use the fraction of diffuse to direct radiation to tune away the direct effect of misrepresenting rain events duration on GPP.

We added a discussion of the limitations of our study and potential generalization of our results in the previous summary section (section 4, now renamed Summary and discussion). This discussion on light use efficiency can be found on lines 404-414 in the revised version.

5. JSBACH employs a soil representation unlike most other land-surface models, with the notable difference that it assumes a direct dependence of soil respiration on precipitation, while most other land surface models tend to assume that the moisture control on decomposition acts as a function of soil moisture (and therefore a quantity that integrates rainfall, evapotranspiration, surface runoff and drainage). YASSO's precipitation response has originally been devised to account for long-term spatial differences in moisture and not to represent the impact on rainfall variability on Rh. It is unlikely that many land surface model would exhibit a similar rainfall sensitivity

Our paper is a modeling study and, indeed, we cannot rule out at the outset that the dependency of Rh on precipitation may be model specific. However, we disagree with the statement that it is unlikely that many land surface model would exhibit a similar rainfall sensitivity because soil moisture and precipitation are strongly linked. As shown in the original manuscript on Fig. 3, the decrease in precipitation is accompanied with a decrease in soil moisture. Hence, even in a land surface model where the moisture control on decomposition would be via soil moisture, and not precipitation, we would expect a similar sensitivity. Moreover, Jungandreas et al. (2023) showed that explicitly resolving convection and the associated more intense storm leads to drier soil moisture conditions in their simulations due to higher runoff. Hence, one could even hypothesize that the sensitivity could be bigger in a soil model where the decomposition is tight to soil moisture, as both the reduction in mean precipitation and the reduction in the amount of precipitation going into the soil with explicit convection would act to reduce soil moisture. As such, we believe that our results are relevant for other models and it would be great of course if our study would motivate other land surface model groups to conduct similar experiments. We added this discussion in section 4, which now both summarizes and discusses the results, see lines 418-422.

The method description contains a number of inaccuracies that need to be resolved. The authors claim they use MPI-ESM-LR for their coarse resolution simulation. However, the output is presented not on the spherical grid of MPI-ESM-LR, but on a triangular grid of ICON. Either, the input forcing was regridded to the triangular grid (which is missing from the methods description, but is relevant because this will tend to spatially smear out the native climate model output), or a different climate model has been used to generate the forcing.

The atmospheric forcing for the LR experiment was indeed obtained from the MPI-ESM1.2-LR simulation. However, before it was used to drive JSBACH, the MPI-ESM-LR output was remapped from its native atmospheric grid to the triangular ICON grid used for the 160-km JSBACH simulation. We agree that this preprocessing step should have been explicitly described and we added it now in section 2.2 on lines 105-109.

Minor comments:

L19, but note Georgievski & Hagemann (2019), who show stronger effects of land-use uncertainty.

We revised the corresponding paragraph in the introduction, see especially lines 28-29, to provide a more balanced overview of the various uncertainties.

L29: This is not fully correct, as many of the studies quoted are either driven by observed climate or re-analysis, but not Earth system model simulations.

This is true, thank for catching this inaccuracy. We corrected the sentence, see lines…

L68: Please stick to commonly accepted terms. The carbon cycle is a biogeochemical cycle and not a biogeophysical process.

Sorry yes, we corrected and also checked the text for other inaccuracies.

L98: This is not correct, amonst others, one model is discretised using a spherical grid whereas the other employs a icosahedral-triangular grid.

As mentioned in one of the responses above to the main comments, both SRM and LR simulations are integrated on the icosahedral-triangular grid, although (and indeed) the forcing of the SRM was also on a icosahedral-triangular grid whereas it was on a spherical grid in LR. We clarified in section 2.2, see lines 38-40.

Section 2.1 describes a version of JSBACH that differs from the version referred to in Section 2.2. This is confusing, because the version referred to in Section 2.2. does not consider PFT or carbon processes

We agree that this can be confusing. We rewrote part of section 2.2 to better explain what version of JSBACH is used in which model version and why we used a certain model version, see especially lines 100-102 and lines 111-114.

L104-105: (a) it is inconsistent to use 1850s climate state together with 1990s CO2 concentration and a justification for this set-up is not given. (b) The statement that the carbon balance between land and atmosphere is closed is misleading because in the simulations presented, there is no interactive carbon cycle between land and atmosphere and therefore the carbon balance is not closed at all. The atmospheric concentration is prescribed and does not react to changes in the land carbon balance. This set-up is the opposite of presenting a closed carbon cycle.

1) Regarding the CO_2 value: as it was the first time that we conducted land surface simulations with carbon at km scale, we wanted to be able to broadly compare the simulated carbon cycle to observations and other projects (like MsTMIPs). The problem is that they all use different CO_2 concentrations (yearly observed values, pre-industrial, present-day…) so we decided to take a somewhat default value representing a late 20th-century, present-day baseline. Using another value for the CO_2 concentration in our offline simulations would have changed the size of the carbon pools, in both simulations, but not the documented differences between LR and SRM, as we could fully explained these differences by differences in the prescribed atmospheric forcing. We will add these considerations in the method section of the revised manuscript. Also note that in the revised version, we don't use 1850 climate anymore (see response to reviewer comment 2).

2) Regarding the 'closed' carbon balance: We agree. Our intended meaning was that the terrestrial carbon pools were spun up to a steady state (i.e., equilibrium) where the net land-atmosphere flux is zero. We don't use closed anymore.

L171: correct, but why do you then not investigate the alternative differences stemming from the different radiation schemes?

This is technically not feasible. It would require implementing either the SRM radiation scheme in the LR model or vice versa, which cannot be justified for the scope of this study.

We believe that the evidence provided, also taking into account the extensive supplementary analysis conducted for this review, is convincing. Furthermore, as noted in the literature (e.g., Prein et al., 2015; Vergara-Temprado et al., 2020), when transitioning across the convective 'grey zone' to km-scale resolutions, the shift from parameterized to explicitly resolved convection dominates the differences in surface shortwave radiation, far outweighing the mathematical differences between modern radiation schemes. We will revise the corresponding lines to clarify.

Reference

Chakraborty, T. and X. Lee, 2021: Large differences in diffuse solar radiation among current-generation reanalysis and satellite-derived products. J. Climate, 34, 6624-6650.

Jungandreas, L., C. Hohenegger and M. Claussen, 2023: How does the explicit treatment of convection alter the precipitation-soil hydrology interaction in the mid-Holocene African humid period? Clim. Past, 19, 637-664.

Zhang, Y., Bastos, A., Maignan, F., Goll, D., Boucher, O., Li, L., Cescatti, A., Vuichard, N., Chen, X., Ammann, C., et al.: Modeling the impacts of diffuse light fraction on photosynthesis in ORCHIDEE (v5453) land surface model, Geoscientific Model Development Discussions, 2020, 1–35, 2020.