Response to Reviewer 1

Overall, this is an interesting study. The idea that precipitation duration in km-scale climate forcing may affect shortwave radiation and land carbon fluxes is worth exploring. However, the manuscript still needs substantial additional support. At this stage, many results are based on comparisons between two model setups, but it is still not clear whether the SRM forcing is more realistic, or how much of the GPP difference can really be attributed to the treatment of convection.

We sincerely thank the reviewer for the time and constructive evaluation of our manuscript. The comments identified key areas requiring stronger attribution and methodological clarification. As detailed in our point-by-point responses below and to be able to adequately answer the reviewer comments, we conducted significant new analyses, putting our results into context with observations. We also conducted several new, targeted offline sensitivity experiments to strengthen the previous attribution of GPP differences.

While working on the revisions, we realized that we had made a suboptimal decision on the original set-up. In the submitted version, we had compared multi-year "2021" climate from SRM with multi-year piControl climate from LR. At the time of the study, the multi-year "2021" 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). In hindsight, this was an unfortunate decision as we are comparing two different climates. For this revision, we simulated again LR taking the atmospheric forcing from that same MPI-ESM-LR model, but this time from the year 2021 of the ssp2-4.5 scenario, as that year has about the same CO_2 concentration in the atmosphere as the SRM (414.4 ppm versus 414.39 pm). 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 probability density function of rain affects GPP and explicitly resolving convection enhances GPP under strong precipitation (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; 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 which we attributed 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 due to the drier conditions. Given the sensitivity to the simulated climate, we then use the comparison to observations to put the results into context, as suggested by the reviewer. All in all, and incidentally by having simulated two distinct climates with the LR model, 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.

1. The paragraph around L50, where the authors explain that the two forcing datasets differ in many aspects beyond convection, is important. However, it may fit better in the Methods or Discussion rather than in the Introduction. The two forcing datasets differ in more aspects than just the treatment of convection. However, acknowledging this limitation does not resolve the uncertainty in the interpretation. Since SRM and LR differ in model resolution, climate model, forcing frequency, land-surface representation, and background climate, it remains difficult to determine how much of the GPP difference is actually caused by explicitly resolved convection or shorter rainfall duration. The additional sensitivity tests should be provided to better support this attribution.

We agree that acknowledging the structural differences between the two models in the Introduction interrupts the flow and weakens the premise before the results are presented. Following the reviewer's advice, we relocated this discussion to the Methods section, see last paragraph of section 2.2.

Concerning the fact that the two forcing datasets differ more than by the treatment of convection, we have conducted additional simulations to test the effects of potential confounding factors. In a nutshell, they confirmed the main result of the paper, i.e. that the use of a convective parameterization leads to an underestimation of GPP under strong precipitation events due to too little shortwave radiation associated with long lasting events (for the same amount of precipitation).

First, as explained above, we conducted a new LR simulation with a different background climate. It is now the more correct climate ("2021"), but at the same time, incidentally, comparing these new results to the previous results allows us to more definitely isolate the effect of the background climate. Differences in the mean GPP-precipitation relationship for the strong precipitation events do not depend on the background climate (see Fig. R1.1 below).

Second, to further and more unequivocally tie the differences in the mean GPP-precipitation relationship to the treatment of convection, we used the baseline LR configuration (from our new 2021 climate) but selectively substituted its atmospheric forcing by the one taken from the SRM configuration: (a) precipitation (Pr), shortwave radiation (SW), and air temperature (Ta); (b) Pr and SW only; and (c) SW only. As shown in the Fig. R1.1 below, replacing SW only is sufficient to remove the decline in GPP with intensifying precipitation. This explicitly isolates the meteorological forcing as the dominant driver of the GPP increase by high precipitation amounts in the SRM, demonstrating that the differences are not due to differences in model resolution or land surface presentation. Moreover, as for these sensitivity experiments, we downgraded the 3-hourly frequency of the SRM atmospheric forcing to the 6-hourly frequency of the LR, the difference in forcing frequency can also not explain the distinct sensitivity of GPP to precipitation. Figure R1.1 also reveals that replacing SW only, even though it removes the decline of GPP, leads to much too large GPP values. Replacing precipitation and temperature is needed to bring the GPP values in LR close to the SRM ones, with a larger impact of precipitation. As noted in the submitted version, the two forcing datasets simulate different meteorological conditions, and this modulates the expected GPP increase from explicitly resolving the short-lived nature of storms in SRM.

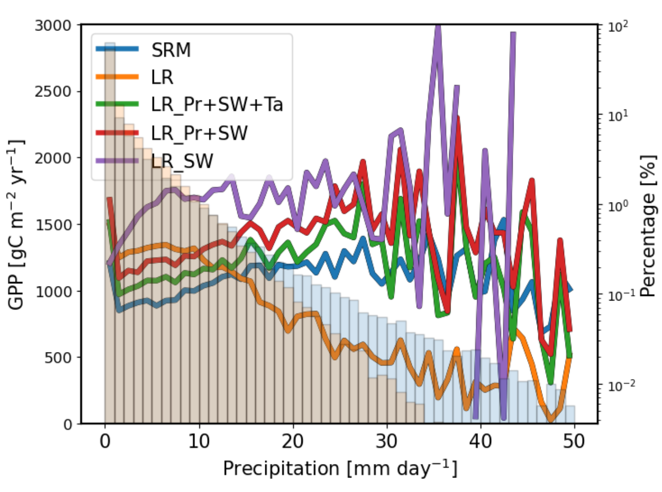

Fig. R1.1: *Daily mean GPP as a function of daily mean precipitation for LR (orange line), SRM (blue line), and three LR sensitivity experiments in which selected atmospheric forcing variables are replaced by their SRM counterparts: precipitation (Pr), downward shortwave radiation (SW), and air temperature (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.*

We changed the LR baseline simulation to the new ("2021" climate) LR simulation and correspondingly updated Tables, Figures and text when necessary. We also mentioned the results of the additional sensitivity tests in section 3.2.1 on lines 217-225 and added Fig. R1.1 in the supplementary information as Figure S1.

2.The reason for fixing atmospheric CO_2 at 367 ppm in both simulations is not sufficiently clear. This choice may be reasonable, but the motivation should be explained more directly.

We chose this specific setup for three reasons. First, we use a constant value so that we can look at the equilibrium response, which is easier to interpret than a transient climate. Moreover, it also requires less integration years, which is of relevance given the computational cost of the 5-km simulation. For instance, a 30-y historical simulation would require 30 years plus some members to sample internal variability, whereas here we only need to simulate a couple of times the same year to sample variability. Second, we need to use the same value in both simulations, otherwise differences in carbon pools could be due to the different atmospheric concentration in the land model. Finally, 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 added these considerations in the method section, see lines 119-124.

3.The Results section mainly compares LR and SRM, and most of the results show differences between LR and SRM. This is useful, but it does not really tell us whether SRM is better or more realistic. The paper needs more quantitative evaluation, not only maps and mean differences between the two simulations. Otherwise, the results show that the two simulations are different, but not necessarily that the SRM forcing improves the land carbon simulation. The paper mentions FLUXNET, MODIS, and MsTMIP, but this is mostly a broad comparison of carbon flux magnitudes. Why not directly compare the key variables with observations or high-resolution products?

We are in sympathy with the reviewer comment. The main goal of the paper was not to determine whether the SRM is better or more realistic, but how the representation of the distribution of precipitation rates affects GPP. The broad comparisons to Fluxnet2015, MODIS, and MsTMIP at the end of Section 3.1 were included solely as a 'sanity check' to demonstrate that our simulated fluxes operate within a biologically reasonable parameter space for the European domain, rather than as a strict validation metric.

Having said this, and given the focus of the paper on the GPP-precipitation relationship and also given the role of the simulated mean climate for explaining the domain-mean GPP difference between LR and SRM, we agree that we should have used observations to contextualize these two aspects, which we did.

First, Figure R1.2a below shows the GPP precipitation relationship derived from FLUXNET station data. From the FLUXNET dataset, we only kept stations with at least one full year of non-missing GPP observations between 2018-2022 and we keep one year per station to avoid giving more weight to specific stations. We choose five years as a compromise between enough stations and not a too strong climate change signal, but the results are not dependent upon the chosen number of years. This gives us 118 stations. The observations clearly show that GPP does not decrease once it starts precipitating. There might be a very weak decrease in GPP between 15 and 25 mm day^{-1} but in no comparison with the decrease seen in LR. As such, the observations confirm the relationship as simulated by the SRM. For completeness, we note that in the SRM, GPP increases between 15 and 25 mm day^{-1} of precipitation, an increase not seen in the observations. To further confirm the results, Figure R1.2b also shows surface shortwave radiation as a function of precipitation as it is the driver for the changes in GPP under high precipitation amounts. The observations are taken from the ECAD-EOBS dataset. The latter is a high-quality gridded product based on station observations, available only over Europe, widely used to validate regional simulations conducted over Europe. Corroborating Fig. R1.2, we see only a weak decline in SW radiation of about 7 W m^{-2} between 10 and 20 mm day^{-1}, compared to 50 W m^{-2} in LR. Hence, the observations confirm that SRM can more realistically capture the relationship between radiation and precipitation, avoiding a strong decrease of radiation by high precipitation amounts, leading to a more realistic GPP-precipitation relationship.

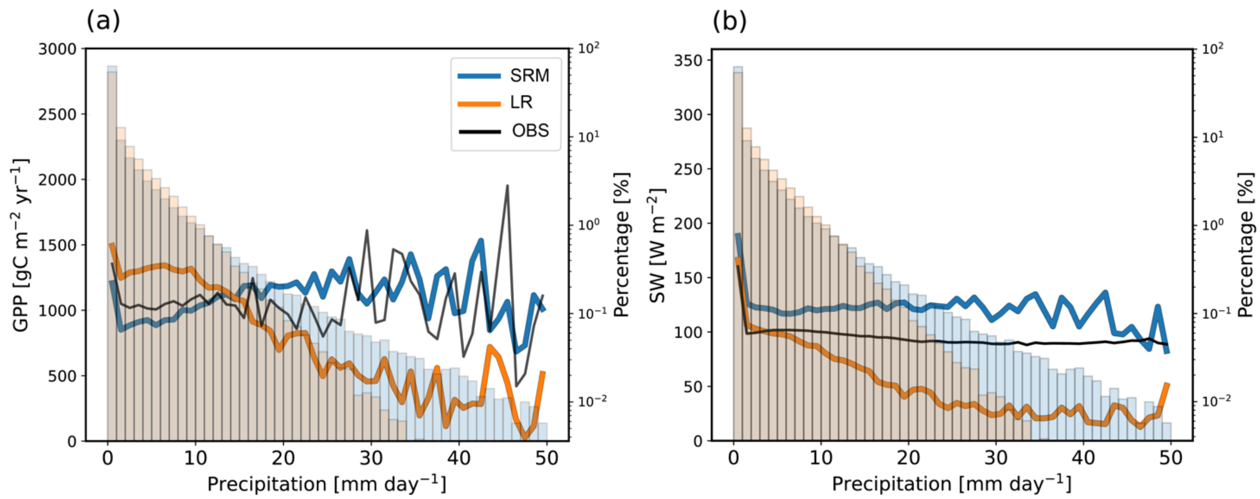

Fig. R1.2: ***(a)*** *Daily mean GPP as a function of daily mean precipitation for LR (orange line) and SRM (blue line) together with their histogram of precipitation (bar). Black line for observations, taken from the Fluxnet dataset and retaining one year of observations per station with at least one full year of non-missing values between 2018-2022 (118 stations).* ***(b)*** *Same as (a) but for surface downward shortwave radiation. Observations are taken from the gridded ECAD-EOBS dataset and averaged over the years 2018-2022. We consider the mean over the years 2018-2022, to capture some of the internal variability without being too much affected by the climate change signal.*

Second, Figure R1.3 below shows a comparison of simulated mean temperature and precipitation to observations given the importance of the simulate mean climate to determine the domain mean GPP. At a first glance, the simulations perform comparatively. Domain-mean differences are -0.6 K in LR versus -1 K in SRM for temperature, and 0.24 mm day^{-1} in LR versus 0.46 mm day^{-1} in SRM for precipitation. The smaller mean error in LR is a result of a better compensation between positive and negative biases, whereas SRM tends to exhibit more widespread positive biases. Over 60% of the domain (see Fig. R1.3a-d), the biases are actually smaller in SRM. In terms of domain-mean GPP, we attributed the differences between the two simulations to larger mean precipitation in the LR. Over central Europe (see Fig. R1.3c-d), LR actually overestimates precipitation, speaking for the SRM being the “more realistic” model. Yet over Eastern Europe, where we observed a strong decrease in GPP in SRM compared to LR due to much drier conditions, the SRM does get too dry over that region during summertime compared to observations. Hence, in terms of the spatial distribution of the climate (and GPP), the performance of both models is similar and which model performs better depends on the selected region.

We added the comparison to observations in the revised version as a new paragraph at the end of section 3.2.1. We also added the observational curve on the GPP-precipitation (Fig. 2a) and SW-precipitation (Fig. 2b) relationship.

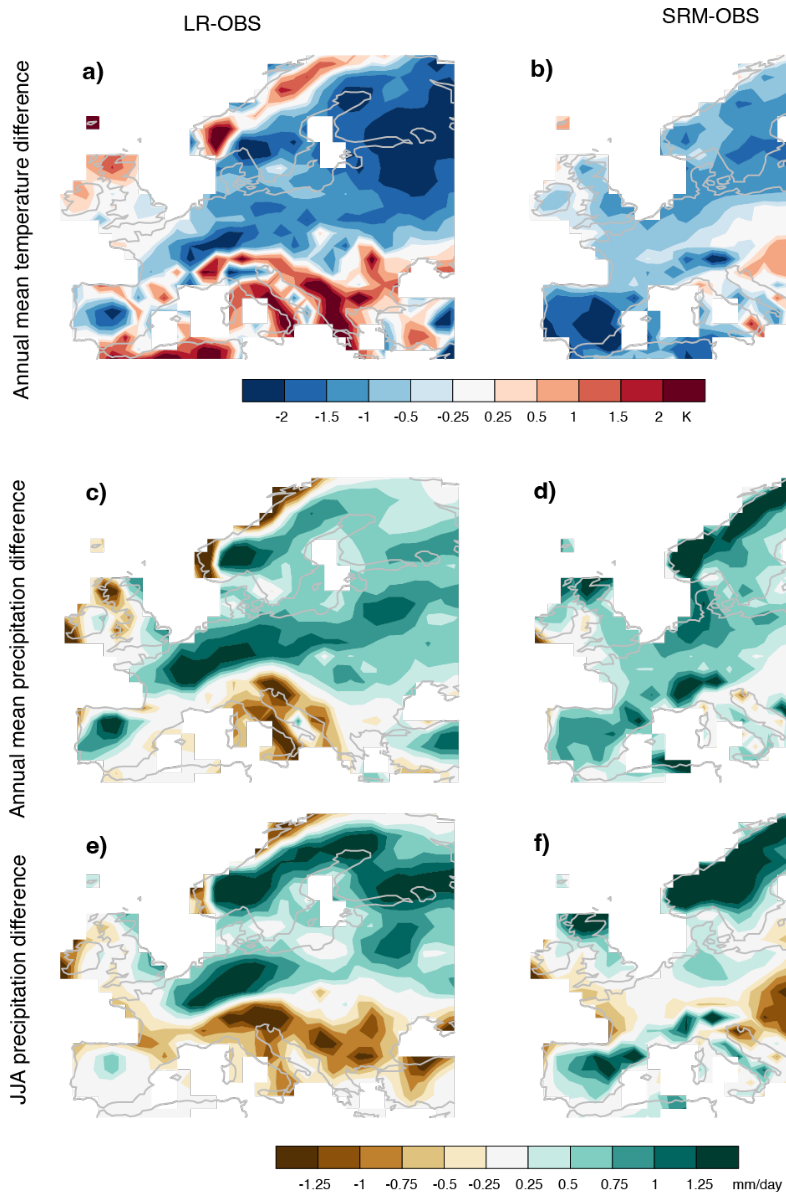

Fig. R1.3: *Comparison of simulated temperature and precipitation to observations taken from the ECAD-EOBS dataset (daily value, original resolution 0.25º). For the observations, we consider the mean over the years 2018-2022, to capture some of the internal variability without being too much affected by the climate change signal. All values were regridded on the coarse-resolution grid of LR.*

4. The increase in domain-mean GPP in SRM is quite small compared with the mean GPP value. Given that SRM and LR differ in many aspects, it is difficult to know whether this small difference really comes from the mechanism highlighted in the paper or from other differences between the two setups. A more direct test would be to focus on the specific days or events that are most relevant to the hypothesis and compare precipitation duration, shortwave radiation, and GPP against observations or high-resolution products.

This would help show whether SRM actually captures the observed rainy-day relationship between rainfall, radiation, and ecosystem productivity better than LR. Without this kind of event-based evaluation, the small mean GPP difference is hard to interpret.

The reviewer asks for event-based observational validation to investigate whether SRM actually captures the observed rain-day relationship between rainfall, radiation and ecosystem productivity better than LR. We now looked at the relationship between precipitation, radiation and GPP in observations, see our response to the previous major comment (Major Comment 3) and the Figure R1.2. However, we did this not on an event basis, as it can be difficult to generalize from single events and there is the danger of cherry picking, but preferred to work directly on the full record.

5. There seems to be a numerical inconsistency in Section 3.2.2. The text says that the mean Ra value is 479.4 g C m^{-2} yr^{-1} in SRM, but Table 1 shows that 479.4 is the LR value, while the SRM value is 519.2. This should be corrected.

The value in the text was indeed a typographical error.

6. In Summary section, the statement that the land surface in SRM "takes up more carbon" should be clarified.

This does not apply anymore as we adapted the text to fit the results obtained with the new LR baseline simulation.