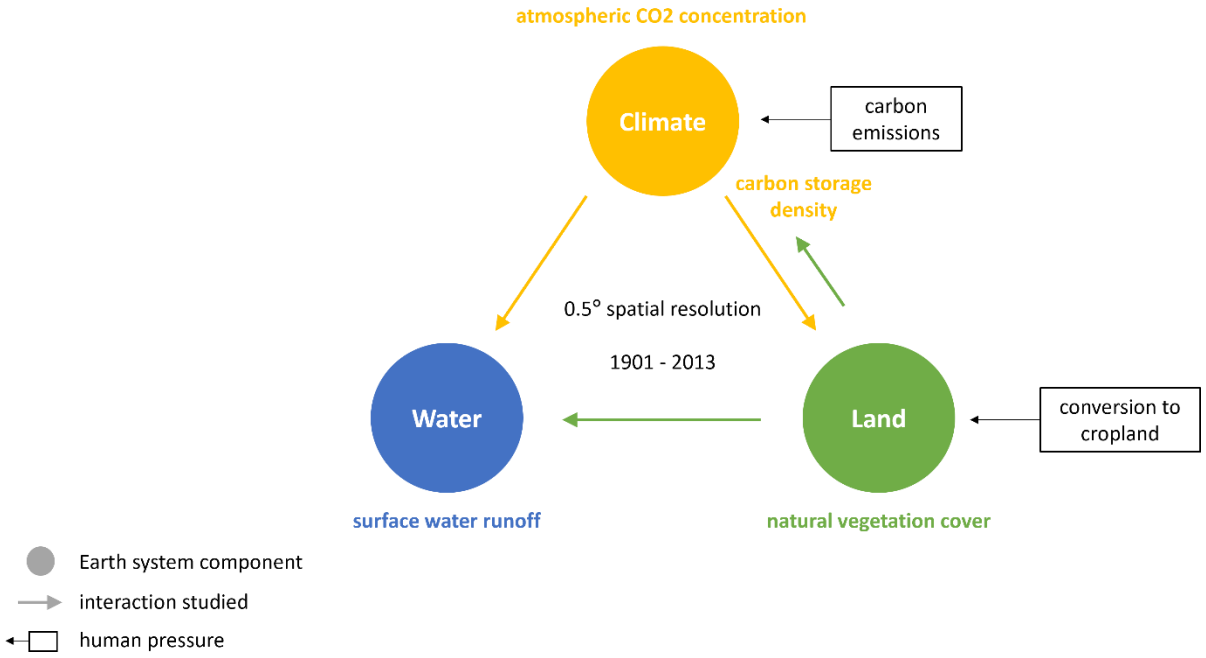
The fundamental question, which might highlight my lack of understanding, is that the authors do not actually do this [*communicating the spatially resolved global pattern of crucial Earth System interactions, ed. note*] according to what I understand to be Earth System interactions. […] this paper is about "Earth System interactions" but LPJmL is not coupled. It's forced by a reanalysis. In my naïve thinking about Earth System Interactions, I think of a coupled system where LPJmL is coupled to (say) ECEarth. I think of something like the GLACE experiments undertaken and published by Koster et al. We know that land models behave very differently once coupled. It is inconceivable to me that this would not be true of LPJmL. So, how robust are the results to being run offline and uncoupled? Might this affect the conclusions?

*As correctly pointed out, we are deriving Earth system interactions from LPJmL being driven by reanalysis data, which means that atmospheric processes are driving terrestrial processes, but not the other way round (**one-way coupling**). In order to assess the interaction strength land -> climate, we are drawing on vegetation carbon, i.e. the density of carbon being stored in plants, as a proxy for the climate. Strengths of feedback from the water cycle to atmospheric processes are not being considered in this study. The offline approach allows us to neatly isolate particular effects without the need to disentangle overlaying feedback effects. We will clarify our understanding of an interaction in the manuscript.*

Furthermore, in response to both your comment and Comment 2 by Reviewer #2, we will add a conceptual diagram, visualizing the set of Earth system components and interactions studied (see first draft below).



They do not tell me anything about this reanalysis, only referring to a paper by Lade et al. (2021). I accept that if this paper fully explains the reanalysis, then it does not have to be fully re-explained. But the paper I am reviewing here says nothing except the spatial resolution. It does not even name the reanalysis, I think.

Thank you for pointing out this lack of clarity. We will add some basic information on the reanalysis data and furthermore, reference Harris et al. 2014, who introduced the dataset.

Fundamental to reanalyses are data that form the reanalysis and that cannot conceivably be equally thorough across the whole world. Does this matter? How sensitive are the results presented here to uncertainty in the reanalysis? Looking at Figure 3 I would suggest that a lot of the red in Figure 3b, and almost all the patterns in Figure 3c and 3d are located in places where the reanalysis would be least

reliable. So, does this matter? If the reanalyses are perturbed by (say) +/- one standard deviation how do the patterns change? I am not saying the results are wrong, but I am saying that I cannot tell if the results are robust and that is a problem for a reviewer.
The reliability of reanalysis-datasets like CRU TS indeed shows a spatial variation, reflecting the density of meteorological stations. While this uncertainty might translate to the interaction strength of a ***single*** *tile, we would like to point out that our main results (varying patterns in the performance of natural Earth partitions, interaction profiles of bottom-up clusters) are based on* ***large-scale*** *trends. We will add these reflections to the manuscript.*
Nevertheless, we agree that more transparency about the spatial variability in reliability is needed in the manuscript. CRU TS allows for an objective assessment of the reliability of particular data in the form of spatially and temporally resolved station counts and station influences (Harris et al. 2014). We will add a paragraph to the Discussion, describing which areas are particularly data-sparse during the time frame relevant for this study and, consequently, which results should be interpreted with caution.

Second, and understandably, the authors only use LPJmL. This is a highly respected model, but it is only one model. I am not suggesting that the authors need to repeat this with other dynamic vegetation models but how robust are the results to minor changes in LPJmL? There are plenty of examples in the literature that suggest LPJmL is a good model, but fairly there are also plenty that suggest it has its issues like any global model does. So, do the maps and patterns change if elements of the model are modified? Of course, you cannot do a full analysis of this, but some effort to determine which of the results are robust to some of the uncertainty in LPJmL feels warranted.
We agree that there might be a certain degree of variability both across global vegetation models and across model settings in LPJmL. However, it is difficult to quantify a degree of "uncertainty" when corresponding ground truth data (e.g., in the form of equally highly resolved observational data) is not available.
*Nevertheless, in alignment with Lade et al. 2021, we are happy to conduct a sensitivity analysis with respect to LPJmL being driven by different climate models (*CanESM2, GFDL-CM3, GFDL-ESM2G, HadGEM2-AO, MIROC-ESM, NorESM1-ME*). The data of the corresponding model runs is publicly available and could demonstrate a general robustness of our main results with respect to slight deviations in the input data.*

Page 3 why is the aggregation step omitted and does it matter?
Lade et al. aggregate tile-wise simulation outcomes by continent and vegetation zone ***before*** *estimating interaction strength. In doing so, there is a risk of "blurring" local interaction patterns. In contrast, we estimate interaction strength for every single tile to capture these very local patterns. In a second step,* ***after*** *estimating interaction strength, we evaluate different aggregations based on how well they represent the local patterns.*

Page 9, line 187. I doubt this explanation is true – I suspect it's linked to an increase in evaporation.
On page 9, l. 187, we hypothesize that the positive effect of land use change on surface water runoff in central Australia, southern Africa, and the periphery of the Sahara desert is likely due to the higher rainfall infiltration rate of cropland in comparison to the former barren soil. While this might be one factor, we agree that the increased evapotranspiration might be another important factor. The latter is being supported by the results in Sterling et al. 2012. We will add this hypothesis and the reference.

Page 10, line 213 – I thought this was regionally specific?
On page 10, l. 213 we state that the observation of positive interaction strength between climate change and surface water runoff in many tropical forests aligns with Zhou et al.'s (2013) hypothesis that in the regions where vegetation cover is already close to saturation, an increasing level of atmospheric CO2 mainly leads to a decrease in transpiration and thereby to an increase in runoff. Recall that for the quantification of the effects of climate change on surface water runoff and on

*vegetation cover, we use **global** atmospheric carbon dioxide concentrations, as being explained in the Methods section. By quantifying the effects of change in this global variable on the **local** values of runoff density and vegetation cover, we reive a **local** interaction strength.*

Finally, all the conclusions might be 100% right, but I cannot tell. I cannot determine if they are broadly real, or merely the consequence of using one model with one reanalysis. I am not sure this meets the criteria for a publication in a significant journal.
[…] I honestly cannot determine whether the results are artefacts of the reanalysis, or of LPJmL, or of the other techniques in the paper. It feels very much like "trust me" and I try to be sceptical when reviewing papers. So, at least for me, this paper needs a fundamental re-write to *not* merely present results, but to provide rigour to demonstrate that the results are sound and can be interpreted beyond this specific modelling system.
Thank you very much for raising this point. As outlined further above, the fact that our main results are based on large-scale trends naturally entails a certain robustness with respect to deviations in the input data. Nevertheless, we agree that both a discussion of the spatial variability in reliability and a sensitivity analysis with respect to model settings in LPJmL will enhance the manuscript. Concerning the later, we will repeat our analysis for six runs of LPJmL, each driven by a different climate model.