Response to Referees

We thank both reviewers for reading our manuscript and providing constructive criticism. We are pleased with the recommendations for publication from the reviewers: referee#1 *"In general, the paper is written and structured well and fits into the scope of the journal."* and referee#2 *"I recommend the publication of the manuscript after minor revisions."*

In the following we address each of the comments of the reviewers point-by-point. For that we paste the original comments in red and our response in black. Our corresponding changes to the manuscript are presented cursive and as separate blocks.

Referee#1

The authors evaluate the performance of the new JSBACHv4 biosphere model in comparison with the JSBACHv3 model in the context of the new ESM model developed at MPI Hamburg. In this paper they have a focus on fast (sec to year timescale) processes. In general, the paper is written and structured well and fits into the scope of the journal. Because of the focus on evaluation than on model description the paper lacks mathematical description of the model changes. Perhaps adding some formulae would be beneficial. But I see the constraints due to the length of the paper.

Indeed our focus is on assessment and not on model description as the scientific contents of the new JSBACHv4 are a subset of JSBACHv3 and that model is comprehensively documented (Reick et al, 2021) with all mathematical detail. We emphasize this better in the revised version by adding "(documented in Reick et al., 2021)" in the following sentence (line 100):

"However, it applies the same parametrizations as JSBACHv3 (documented in Reick et al., 2021) and includes the additional feature of frozen soil water and a five-layer snow scheme (Ekici et al., 2014; de Vrese et al., 2021).“

*C.H. Reick et al. (2021), JSBACH 3 - The land component of the MPI Earth System Model, documentation of version 3.2, Reports on Earth System Science, 240, https://doi.org/10.17617/2.3279802.

Specific comments:

The model setup has to be described in more detail. The model setup is distributed at different parts in the manuscript (e.g. prescribed ocean SST, prescribed PFT distributions).

Thank you for noting this. Indeed, it is not very fortunate that the description of the model setup spreads over different sections. We reordered the former sections 2.1 and 2.2 and added a new section "Simulation setup" to the methods. Accordingly we had to adapt the former text. As the new section (2.2. Simulation setup) encompasses a considerable amount of text, we refrain from quoting it here and suggest to read it in the newly submitted manuscript.

NPP has been selected for comparison with observational data. It would be interesting to see above-ground biomass as an additional validation data set.

The main distinction between the carbon pools representing plant organs in JSBACH is by turnover time, namely fast and slow compartments. Accordingly, fine roots and leaves are combined in a single carbon pool, and stem and coarse roots also in a common carbon pool. Therefore in the carbon pool model above and below ground carbon are not distinguished and thus, although interesting, the suggested comparison is not feasible for our model.

Due to the prescribed vegetation distribution cover cannot be used as an evaluation. In general is would be interesting to the see the performance of the ESM for a dynamic vegetation without prescribed PFTs. But this would be a topic for a separate publication.

We agree.

The formatting of tables could be improved. The vertical lines in the headers do not correspond. In particular a separator for the two JSBACH versions in the header is missing.

We agree and re-formatted all tables accordingly.

To quantify the bias between simulation and observational data the normalize mean error (NEM) metric might additionally be used (Kelly et al, 2013).

A good suggestion. We calculated the Normalized Mean Error for all assessment variables of the two models (JSBACHv4 and JSBACHv3) and show them in two tables:

Table A1. Normalized Mean Error (NME) of JSBACHv4 relative to observations.

	Albedo VIS	Albedo NIR	LST	LAI	FAPAR	NPP	WUE
Year	-	-	0.138	1.294	0.796	0.783	1.341
January	0.267	0.410	0.149	1.266	0.604	0.569	-
July	0.411	0.508	0.193	1.102	0.757	0.777	-

Table A2. Normalized Mean Error (NME) of JSBACHv3 relative to observations.

	Albedo VIS	Albedo NIR	LST	LAI	FAPAR	NPP	WUE
Year	-	-	0.147	1.430	0.669	0.658	1.406
January	0.232	0.415	0.164	1.304	0.687	0.576	-
July	0.400	0.452	0.176	1.310	0.688	0.653	-

However, as proposed by referee#2 we also added a Taylor Diagram that visualizes the model performance even better. Therefore we added the tables to the appendix and only mention its results in the new “General Performance” section (see answer to the referee#2).

Referee#2

The authors evaluate the performance of JSBACHv4 within the ICON ESM and JSBACHv3 within the MPI ESM versus a set of observational data. The authors rely on coupled land-surface-atmosphere simulations with prescribed sea surface temperature and sea ice. Variables evaluated include albedo, Land Surface Temperature (LST), Terrestrial Water Storage (TWS), Leaf Area Index (LAI), Fraction of Absorbed Photosynthetic Active Radiation (FAPAR), Net Primary Production (NPP), and Water-Use-Efficiency (WUE). Biases between model results and observations are substantial in many variables. JSBACHv4 performs similarly to JSBACHv3 (line 604) as the process description are almost identical for both models. This raises the question of why JSBACHv4 was not improved relative to JSBACHv3 to avoid some of the major model biases.

The transition from MPI-ESM to ICON-ESM is an ongoing major effort at our home institution MPI for Meteorology with a history of more than one decade of technical and scientific development. So far the development of ICON-ESM concentrated separately on the major Earth system components (atmosphere, ocean, sea ice, land). Even if the components work well when run in isolation, the combined model typically performs less well because of the unforeseeable effects of the additional degrees of freedom arising from the interactions between the components. This makes it necessary to further tune the combined model, and this is another major effort. Tuning concentrated first on the stability of ocean circulation and the reproduction of historical mean climate. The resulting ICON-ESM Version 1.0 is a major milestone of our model development documented in Jungclaus et al. (2022). Next tuning steps for various other aspects must follow. Even a well equipped institute like ours cannot maintain in parallel two ESMs for years. Therefore research using our new ICON-ESM scientifically started early, well knowing that its development is not finished, but also an unfinished model can be used for serious science when being aware of its particular deficiencies. This is one major reason why the assessment layed down in the submitted manuscript is so important for the many (of in particular the German scientific community) working on land issues with this model. – We hope that this explains why we want to publish this assessment already now and don't wait until biases, that we partly got aware of only because of this assessment, are remedied.

Land carbon cycle and climate-induced biogeographical changes in landcover are not assessed. Modules of JSBACH representing the latter two are switched off. It is a weakness of this study that biogeographical changes and the land carbon cycle modules are switched off and not evaluated as the distribution of plants has an impact on albedo, WUE, etc.

We agree. Nevertheless, we want to point out that for such a major new development as ICON-ESM (see above) the long time scales of biogeographical changes and land carbon turnover stand at the end of the tuning efforts, because first the mean climate must be correct. But unfortunately we are not at this tuning state yet.

I recommend the publication of the manuscript after minor revisions.

We are happy to read this!

Specific comments

1. The authors link biases in albedo and related variables to the applied soil albedo and canopy map (line 326) and fixed minimum and maximum albedo values (line 322) all already used in JSBACHv3. Surprisingly, the authors do not update these features or at least the albedo map in JSBACHv4. This downgrades the manuscript somewhat to a progress report. It seems a necessary next step is to update the albedo module to reduce the large biases in model outcome versus observations. It would make the paper more interesting and useful if these updates in JSBACHv4 would be implemented (maybe this is for computational and personal reasons not possible?). Otherwise, the risk is that this manuscript is outdated very quickly.

We agree but must once more point to the circumstances of model development at MPI: That it will be necessary to update the albedo maps is one of the outcomes of this assessment. We got aware of this only after the many simulations by which the performance of ICON-ESM Version 1.0 is documented in Jungclaus et al. (2022). Moreover, it is this model version that is already scientifically used in our community so that it is important to document these biases.

2. A graphic that summarizes the outcomes of the evaluation concisely is missing. For example, a so-called Taylor diagram should be added to show data-model agreement across variables.

We added a new section "General Performance" to the results where the corresponding Taylor Diagram (see figure below) is shown.

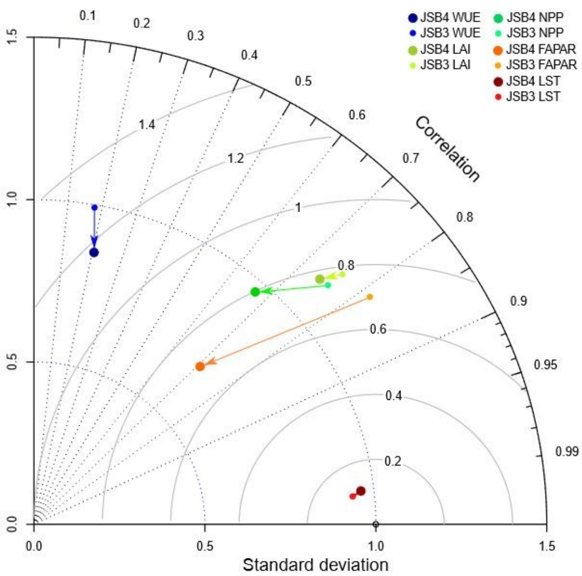

Figure 1. Taylor Diagram (Taylor, 2001) of normalized pattern statistics for annual means of our main evaluation variables LST, LAI, FAPAR, NPP and WUE. The diagram contains no values for albedo because during polar winters observations are missing. JSBACHv3 is shown in small dots and JSBACHv4 in big darker dots. Arrows indicate the change from JSBACHv3 to JSBACHv4. The centered pattern root-mean-square error (RMSE) and standard deviations have been normalized by the observed standard deviation of each field before plotting. Correlations are depicted by lines from the origin representing angles relative to the horizontal base line. Standard deviations are represented as arcs around the origin. Normalized centered pattern RMSE values are seen as circular arcs around the value 1 of the baseline (normalization on standard deviation 1.0).

We added the following text (line 355):

"Taylor (2001) proposed a graphical way to depict the overall performance of a model. The corresponding Taylor Diagram for JSBACHv4 and JSBACHv3 is shown in Fig.1. The LST achieved the by far best agreement of the statistical metrics with observations. These metrics remained nearly unchanged from JSBACHv3 to JSBACHv4. Except for the LST all standard deviations are reduced in JSBACHv4 (especially FAPAR). Overall some statistical metrics of JSBACHv4 improved while others worsened as compared to JSBACHv3. Thus the overall performance remained more or less the same. This is also visible in the NME in Table A1 and A2, where the NME magnitude is very similar between the assessment variables of both models."

3. A display documenting the seasonal evolution of snow cover is missing. Showing only January snow cover is not enough for a proper evaluation.

We show snow cover only as an auxiliary variable to explain the albedo bias, assessment of snow cover itself is not intended. Indeed we could extend our assessment to snow cover, but we decided to concentrate on variables that picture integrated behaviour across many processes, and for albedo snow cover is only one component. Accordingly, we prefer not to add a separate assessment of snow cover. Obviously we did not make this distinction between main assessment variables and auxiliary variables sufficiently clear in the text and add clarifying remarks (line 144):

"To shed some more light on the origin of biases in the selected assessment variables, we consider also some auxiliary variables, these are introduced below jointly with the respective assessment variable."

In addition we updated table 2 and plotted the assessment variables in contrast to the auxiliary variables bold. In each case we think that the interested reader gets at least a first idea of the seasonal course of biases from the two plots for January and July that we show.

4. Section 2.1 and 2.2 Spin-up of the land model is not mentioned. Is no spin-up required when land carbon and biogeographical changes are switched off?

Indeed, since the dynamics of land carbon and biogeography are not active in our simulations, the longest remaining time scales reside in soil memory. Because the AMIP simulations underlying our assessment start from a historical simulation of the related full Earth System Model, the soil water reservoirs are upon start already filled to a realistic level. We address this topic in our new section "Simulation setup" that was included in reaction to a comment by reviewer #1 (see line 131-139):

"In case of ICON-A+JSBACHv4, data from year 1980 were taken for initialization from the respective historical simulation (Jungclaus et al., 2022), while the initialization of ECHAM6+JSBACHv3 is based on the state of the year 1979 of the associated historical CMIP6 simulation (Max Planck Institute for Meteorology, 2020). After model start the atmosphere equilibrates within days. Because land carbon and biogeographical components are not active in our AMIP simulations, the slowest land variable in this setup is soil moisture. By the initialization from a historical simulation, the soil water reservoirs are upon start already filled to a realistic

level. Soil water memory is typically a few months, only in desert regions it lasts up to a year (Hagemann and Stacke, 2015). But soil water memory stems (in our model) mostly from below the root zone (Hagemann and Stacke, 2015) so that it is largely decoupled from the active water cycle at the monthly scale that we consider for our assessment. "

* S. Hagemann and T. Stacke: Impact of the soil hydrology scheme on simulated soil moisture memory, Climate Dynamics, 44, 1731–750, 2015.

Minor comments:

- Please number equations.

Done.

- Around L205: GRACE data are used for evaluation of Terrestrial Water Storage (TWS). However, JSBACH does not include aquifers and their changes may influence changes in TWS from GRACE. The authors normalize the model and GRACE data and compare normalized, climatological month-to-month changes to account for these shortcomings. The authors should discuss the implicit assumption behind their approach and potential shortcomings. E.g., relative changes in aquifer water storage are assumed to have the same magnitude and phasing as TWS. How plausible is this?

We agree to the reviewer that we should be a bit more specific concerning the underlying assumptions. In two additions to the article's text we now point out that only the seasonal cycle of TWS is investigated and not long-term trends.

We added the following paragraph in line 217:

"Here we are mainly interested in the question how JSBACHv4 performs in the context of climate and Earth System modelling. For the fluxes to the atmosphere a correct reproduction of the seasonal cycle in TWS is essential. Thus, only changes of TWS in the course of the year are analysed and not long-term trends. Our comparison is based on the assumption that the additional signal from the hydrology below 10 m that is present in the observational data but not in JSBACH does only negligible contribute to monthly TWS changes. We assume that below 10 m depth the hydrological processes are already so slow that they don't add to the phasing of the overall seasonal signal. This pertains in particular to potential signals from aquifers, whose recharging times are much larger than the monthly time scale considered in our comparison (typically decades to millennia)."

We added the following paragraph in line 239:

"Only the average seasonal changes in TWS are evaluated by this method. The amplitude of TWS variations as well as long-term trends in TWS are not considered. Therefore, a low value of Q_{TWS}

does not necessarily mean that variations in TWS are simulated realistically in a quantitative sense, but it shows that the seasonal phasing of TWS is captured by the model."

- L249: I am a bit puzzled that NPP depends on fire. A more conventional definition is that NPP minus any carbon fluxes to the atmosphere from perturbations such as fire, herbivore grazing, pests, and mortality defines Net Biome Production (NBP). Then, carbon release by fire is not part of NPP as suggested here.

Indeed, without further comments our listing of fire as one of controls of NPP can only be misleading: In contrast to e.g. LAI or temperature, the effect of fire is only indirect and thus we better should not mention fire in the revised manuscript. We deleted fire in line 281.

But just to explain: wildfires indeed play a role for NPP because (as implemented in JSBACH) they reduce leaf carbon, thereby LAI, and thus lead to a reduction of productivity.

- Caption Fig. 1: typo Arctic, Antarctic

Done.

- Fig 7: typo: issignificant

Done.

- L380: typo: (5)-> (Fig. 5)

Done (now line 426).