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.

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.

I recommend the publication of the manuscript after minor revisions.

Specific comments

- 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 Otherwise, the risk is that this manuscript is outdated very quickly.

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

Minor comments:

- Please number equations.

- 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?

- 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.

- Caption Fig. 1: typo Arctic, Antarctic

- Fig 7: typo: issignificant

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