Response to Reviewer #2

Reviewer_2_ Main_Comment_001: The manuscript by Calle and Poulter investigates age-class dynamics as simulated with a dynamic global vegetation model (DGVM) called LPJ-wsl 2.0, a model developed based on the DGVM LPJ. Some aspects of this model are described in the methods, including those which were newly introduced to work with age-classes. The core of the paper seems to be a set of factorial simulations on different spatial scales used to investigate age-class dynamics together with their effect on the simulated carbon fluxes. In addition, the authors assess the contribution of the two types of modelled disturbances (fire vs land use) on forest age structure and derive a generalised linear model to predict carbon fluxes from temperature, precipitation and age-class. The latter is then used to map the "effective range" of each of the predictors to identify regions with significant contribution of demography. I find the manuscript interesting and timely, because forest age structures are an important aspect of the (anthropogenically) disturbed terrestrial biosphere, particularly with respect to the role of land use in climate change mitigation scenarios, and since forest age structures are still underrepresented in DGVMs. In my opinion, however, several aspects of the paper need careful revisions. In particular, the main aim of the paper did not become apparent to me (see general comments below).

Reviewer_2_General_Comment_001: The main aim of the paper is unclear to me and so is what the new aspects are (i.e. the gap mentioned in the abstract l.15 and in the last paragraph of the introduction ~l.81). Is the paper supposed to be a) a model development paper, i.e. describing LPJ-wsl v2.0 or describing the implementation of age-classes in LPJ-wsl v2.0? Or is the paper b) the investigation of the simulated demographic effects? While I find detailed descriptions of models / new model development important and a legitimate scientific contribution, in my opinion, there would still be quite a bit information missing if a) would be the purpose of the paper. To me it especially did not get clear, what has been the new development and what was there before (particularly in subsection 2.2.2 – is this all new or are parts from LPJ-wsl v1.0 or even LPJ?). One aspect that could help to clarify this would be a consequent use of "LPJ-wsl v1.0" vs "LPJ-wsl v2.0" (vs LPJ) highlighting the "modifications for integration with age-classes" (l.118). (Furthermore, there are currently several occurrences of only "LPJ" which probably should be called LPJ-wsl 2.0 (e.g. Table 1, Supplementary, results section)). In addition to having clear model version references, some reordering could help, e.g. moving LPJ-wsl v1.0/LPJ aspects to 2.1.2, such as probably most aspects of fire, primary and secondary/managed forest, LUH2 driver, emissions and residues, product pools, etc. C2 This could, by the way, also solve the sudden occurrence of primary and secondary tiles (l.215) and the unexplained "land use" in Figure 1.

> We outline the main aims of the paper in the first sentence of the last paragraph of the Introduction,
>
> > "The overall aims of this study were to present new model developments that simulate the time-evolution of age-class distributions in a global ecosystem model and to determine if explicit representation of demography influenced ecosystem stocks and fluxes at global scales or at the level of a grid-cell."

> There are two main aims of the paper. The first aim, (a) to present technical model development details. The ageclass developments are the new developments, which we tried to make clear in the title and abstract. Although the Fire and Land Use modules have not been changed, we described them in detail for completeness because these modules are integral modeled disturbances that initiate ageclasses; detail knowledge of these processes is deemed important, especially if it helps readers identify points of improvement.
>
> We agree that the naming conventions used varied and this is confusing. We replaced all instances of model version to "LPJ-wsl v2.0"

From the current structure of the paper I tend to assume that b) is the main purpose / the new aspect. In this case – but to some degree this also holds for case a) – I would expect some form of comparison to observational based data, particularly for the global simulation for which the authors derive the role of demography in the global carbon cycle.

> The second aim of the paper is to (b) demonstrate the effect of ageclass model developments on global scale dynamic vegetation simulations. Aside from showcasing the FIA comparisons to provide confidence that LPJ-wsl v2.0 can reasonably represent forest structure attributes among difference ageclasses, we tried to avoid benchmarking.
>
> Firstly, we demonstrate that ageclass improves structural representation, and that this, in turn affects function, which we show via idealized single-cell simulations. One of the main points of the paper, however, is that the model does not simulate every disturbance. And the disturbances we do simulate (fire, land use change, wood harvest) and need improvement. As such, the results we present underestimates ecosystem ages, and therefore the results underestimate the demographic effect.
>
> It is unclear to us as to how other state variables such as global NPP, GPP, Rh, Fire Flux, etc., would change with a realistic representation of forest ages. A benchmarking effort is beyond the scope of this paper, although we are working toward this end to improve our confidence in the flux estimates.

On one hand, I would expect some kind of comparison of the global simulation with and without age-classes to e.g. a GPP or better AGB dataset to get a feeling for the relevance of the finding of a 40 PgC increase in turnover, and, on the other hand a comparison to a global age map, especially since one of the authors recently published such a map (Poulter et al., 2018). The comparison to a global age map could particularly be instructive to learn where the model fails to reproduce age-structures from the observational based dataset and to discuss why this might be the case (e.g. missing disturbances vs. issues with the fire algorithm or as I expect also issues with the LUH2 data – could be included e.g. in 3.3.2 and 4.1).

> We provide a comparison of ecosystem ages in map form and violin plots of ecosystem ages by continent for the GFAD v1.0 age map (Poulter et al. 2018). Much can be learned, even without benchmarking. We know the model underestimates ages because we lack representation of all types of disturbances, from windthrow to beetle kill to small fires.

FireMIP results (Hanston et al. 2020 GMDD) clearly demonstrate that the GlobFIRM module we use in LPJ-wsl v2.0 underestimates burned area.

Reviewer_2_General_Comment_002: I had some problems with the way the matrix notation is presented. In general, I found the matrix description a good idea, since it quite nicely visualizes what happens upon ageing and particularly which fractions are merged into the next age-class. My critique, however, is that this is not what has been done in the code and that it also does not suit any of the two age-class setups applied in the study (Table 1). I would therefore recommend to clearly state that this is the theoretical idea, which neither suits the applied age-class setups (because they both contain unequal age widths) nor is what has been implemented in the code. Furthermore, I would appreciate a paragraph on how the age tracking is actually realised in the code.

We added text to clarify that the matrix formulation is the theoretical basis for the approach. In our paper, we offer four different ways of explaining the VTFT method: 1) the mathematical theoretical description, 2) a plain-language summary of the method, 3) a visual description of hypothetical examples in figure form, and 4) we provide the full model code. We understand that our approach is hard to translate so we tried four different ways of presenting the same procedure in an effort to reach the most people. As a programmer, I find it always easiest for me to view the actual code to understand the implementation more completely. In case you are interested, the main block of code is freely available (also on Zenodo):
<https://github.com/benpoulter/LPJwsl_v2.0/blob/master/src/tools/ageclass_transition.c>

Reviewer_2_General_Comment_003: The authors state that the simulated age structures are an "upper limit of ageclass distributions" due to not represented disturbances (e.g. l.38, l.593) and that the study overestimates ecosystem age (l.606). However, couldn't the simulated disturbances (fire, harvest and land-use changes) also be too strong in some places? Especially with l.233-239 stating that "deforestation always occurs in the ranking of oldest to youngest age-classes. . . typically resulting in greater land-to-atmosphere fluxes"? Could this lead to too young forests as well as to an overestimation of historical fluxes?

We removed the term 'upper limit' throughout. We added clarifying text as below.

"In some locations, it is possible that our wood harvest priority rules (harvest oldest age-class first) might lead to simulated stand ages that are younger than observed stand ages if other harvest rules were applied in practice, such as preferentially logging forests of intermediate age with a goal of preserving the oldest forests from harvest. We evaluated the age distribution by continent simulated by LPJ-wsl v2.0 to the Global Forest Age Database (GFAD v1.0, Poulter et al. 2018), which is derived from country-level inventory data (SM Figure 11). The comparison shows that the simulated ages are consistently older than the GFAD dataset."

The GlobFIRM fire module clearly underestimates burned area, see results from fire-model intercomparisons from FireMIP (Hanston et al. 2020); we have updated the Hanston et al. 2017 reference to the 2020 paper throughout. The FireMIP results confirms similar findings about GlobFIRM.

For the conclusions drawn in this manuscript, two additional scenarios could maybe be valuable: a) a simulation with deforestation occurring in the ranking from younger to older age-classes and b) a "best guess" simulation using the LUH2 harvest categories "harvest from secondary young forest" and "harvest from secondary mature forest".

> Certainly, there are scenarios where we will want to identify regional logging practices for which these encoded rules can be modified. That would be interesting, especially for a focal analysis on full cycle sustainable harvest practices. For simplicity and to reduce the number of 'moving targets', we chose to stay with the old-to-young harvest assumption.

Another thought: Have the authors considered conducting the fully fledged global simulation also with the unequal bin setup? Would this lead to different results?

> It is possible this would lead to different results, but most likely only in an extreme 'very young world' scenario, or in the 'fire band' latitudinal zones, where the ecosystems are relatively young. That's an interesting point, however, and a case to be made for a more flexible setup, such that the ageclass setup could be flexible to accommodate more frequent disturbances. In any case, the largest differences in ecosystem function (NPP, Rh) between the two ageclass setups (equal/unequal binning) are seen in the youngest ageclasses.

Reviewer_2_General_Comment_004: Literature work: In some of the sections I had the feeling that more references / locating the paper in context of the existing literature would be appropriate.

a) First of all I wondered if LPJ-wsl v2.0 is the model on which the publication of Pugh et al. (2019a) is based? In this case this should in my opinion clearly be stated in the paper.

> Agreed. We added clarifying text as below for transparency.
>
> "Technical details are presented for a module representing age-class dynamics, driven by fire feedbacks, land abandonment and wood harvesting in the LPJ-wsl v2.0 Dynamic Global Vegetation Model (DGVM). Prior versions of LPJ-wsl v2.0 that included early technical developments of the land use change module and the age-class module have already contributed to prior studies (Arneth et a. 2017, Kondo et al. 2018, Pugh et al. 2019a).

b) Looking in the literature for occurrences of LPJ-wsl I found several publications that had at least short model descriptions and I wonder why none of these is referred to in the manuscript (e.g. Poulter et al, 2015; Zhang et al. 2017,2018)?

> We understand the confusion. The unique feature of LPJ-wsl v2.0, including earlier versions of LPJ-wsl (as referenced above), is that it is programmed as a fully modular model. Each module can be run independently using compiler flags. This is slightly different than other DGVM models. We have maintained a practice of preserving old

code (bug-free) and adding modular updates to process representation, such that we can revert to older versions of the code. Poulter et al. 2015 does not present model developments. Modular developments for permafrost and wetland methane by Zhang et al. 2017,2018 are not 'turned on' and do not influence our simulation results. At some point, our goal is to conduct a full factorial experiment with all the modular developments, but this is not the aim of this paper.

c) In section 2.2.1 l.142 when introducing the VTFT approach, the authors point to the paper by Nabel et al. (2019) having a similar independently conceived approach. Indeed, it seems as if many of the aspects described in 2.2.1 are similar to those described in Nabel et al. (2019), including the tracking of fractions per year and the merging process: merging of disturbed areas into the youngest age-class and merging of aging fractions exceeding the width of the age-class into the next age-class. Each with subsequent area-weighted averaging of carbon with the transitioning fractions. To a certain degree similarities seem to also hold for the applied age-class setups. While I truly believe that this approach has been independently conceived, I would still recommend relating to the existing approach, e.g. pointing out similarities and in particular also differences.

We added the following text below to the corresponding Section 2.2.1. We clearly state in the text that the VTFT method is similar to that described in Nabel et al. 2020. "The most novel advancement in this study is a new method of age-class transition modeling, which we call 'vector-tracking of fractional transitions' (VTFT), which improves the computational efficiency of modeling age-classes in global models; this is a similar approach independently conceived by Nabel et al. (2019). ". Their paper only provide a brief description of their method. Their focus appears to be on the implications of different age width binning in age class simulations. We add the following text that draws on their findings.

"The age widths of the age-classes in the *10yr-equalbin* setup correspond to common age widths of classes used in forest inventories; for contrast, JSBACH4 uses a 15-year age width in their equal-bin ageclass setup. Most ageclasses in this setup are represented by a vector of 10 elements, wherein each element represents an aerial fraction for each age-unit (Table1)."

A study by Nabel et al. (2020), using the demographically-enabled JSBACH4 DGVM, found that unequal binning of age widths had lower errors than equal age width binning but the largest reduction in model-observation error was achieved by simply adding more ageclasses at younger ages, regardless of the binning strategy employed.

Reviewer_2_General_Comment_005: I would recommend clearly stating when simulation output is referred to as opposed to observational based data (e.g. l.20, l.32-33).

Edited accordingly.

Reviewer_2_General_Comment_006: It did not become clear to me what exactly is compared in 2.3.2 and 3.1: Are these simulation results from a global simulation? From which? Sage? But if from Sage, why are the FIA data with disturbance, stocking or longing excluded?

Reviewer_2_General_Comment_007: Figure 3 and 4: I would appreciate to also have Figure 4 for the 10-year age-widths, since this is what is used in the global simulations. Also, could for ease of readability maybe all panels with unequal age-widths start with the youngest age-class? Furthermore, it might increase comparability when changing the x-axis to show linearly increasing years instead of the classes and then to place the boxes for the different age-classes at age-class mean ages. This would particularly underline the differences in the NEP dynamics among the different age-class setups. Even more so, if the two age-class setups would be integrated in one plot/panel for each of the depicted variables instead of having separate panels with differing x-axis.

Reviewer_2_General_Comment_008: Is there a recommendation/conclusion on what age-class setup to use based on the studied simulations? I.e. when would a simulation with unequal bins be preferable, when with equal bins or the like?

> We clarified in Section "2.2.1 An age-based model of ecosystems – sub-grid-cell dynamics" as below. We think this provides recommendation to use equal-bin setup for global simulations.
>
> "... The within-ageclass elements are not independent and every within-ageclass element has the same state variables, including the same soil water and light. As such, we only simulate processes at the ageclass level, and the within-ageclass elements are a simple method for a 'smooth' transition between ageclass. In theory, we can simulate processes independently for each within-ageclass element, but this is not practical or necessary. The main benefit for using equal-bin or unequal-bin ageclasses is to independently simulate processes. ..."
>
> ".. The use of equal or unequal age class setups is more than just for reporting purposes. There resources differ between age-classes but not within age-classes, and we limit the model to represent a total of 12 ageclasses. Also, there exists a greater range of forest ages at global scales and the equal age-class setup allows us to independently model resource dynamics (space, light, water availability) for more of the terrestrial surface. If we had chosen the unequal-bin setup for global simulations, we would be independently modeling processes only for the youngest age-classes and we would lose capacity to independently model processes at intermediate and older age-classes."

Specific comments/ Technical corrections: —— - Mixed usage of hyphens: grid cell, grid-cell, gridcell; age-widths, age widths; age class, age-class; land use change -> land-use change; land use transitions -> landuse transitions; land-use -> land use, . . .

> We changed 'grid-cell' to 'grid cell' throughout, except when it was used as a joint adjective.

We changed ‘age-width’ to ‘age width’ throughout, except when it was used as a joint adjective to describe the bins.

We changed ‘land use change’ to ‘land-use change’ throughout.

We changed ‘land use transitions’ to ‘land-use transitions.

We verified that ‘land use’ was used appropriately to describe the use of the land.

Reviewer_2_Technical_Comment_001: l.13 “most global ecosystem models” – consider changing to “many” acknowledging the considerable list in l.55-57.

Edited accordingly.

Reviewer_2_Technical_Comment_002: l.15 Could you specify which gap exactly?

We changed the phrasing as below.

“.. This paper aims to present the technical developments of a computationally-efficient approach for representing age-class dynamics within a global ecosystem model, the LPJ-wsl v2.0 Dynamic Global Vegetation Model, and to determine if explicit representation of demography influenced ecosystem stocks and fluxes at global scales or at the level of a grid-cell. ..”

Reviewer_2_Technical_Comment_003: l.18 Could you maybe make this sentence a bit more precise? Could it also be fractions of an age-class which experience a stand-clearing disturbance? The simulated stand clearing disturbance is fire, and the prescribed ones are harvest and abandonment of agricultural area?

A disturbance can occur on a fraction of an age-class, yes.

We rephrased for clarity as below.

“.. The modeled age-classes are initially created by simulated fire, and prescribed wood harvesting or abandonment of managed land, otherwise aging naturally until an additional disturbance is simulated or prescribed. ..”

Reviewer_2_Technical_Comment_004: l.20 “that patterns of ecosystem function” -> simulated patterns? Patterns resulting in/from model simulations?

We added clarifying text throughout to specify whether a statement refers to simulated or observed data, as in “.. that **simulated** patterns of ecosystem function ..”

Reviewer_2_Technical_Comment_005: l.24 land-use change

Edited accordingly.

Reviewer_2_Technical_Comment_006: l.25 “an additional” –in the sentence before, with regards to fire, only the difference between boreal and tropical lats is given, maybe you could give the absolute effect there, too?

> We removed the wording ‘additional’ and simply stated that “Between simulation years 1860 and 2016, land-use change and land management were responsible..”

Reviewer_2_Technical_Comment_007: l.25 “-21 years in temperate (23N-50N) and tropical latitudes” are these analysed together or independently but experience both the same decrease in age through landuse change and land management?

> Temperate and Tropical latitudes were analyzed separately, see results of the statistical model presented in Table 3. But yes, they experienced the same decrease in age over time as a result of land use change and land management

Reviewer_2_Technical_Comment_008: l.32-33 please clarify what kind of “Spatial datasets are provided for global ecosystem age” simulated ones? Do these stem from the ‘fully-fledged’ LPJ-wsl v2.0 simulations?

> We clarified as below. (bold for emphasis, here only)
>
> “..**Simulated** spatial datasets are provided for global ecosystem age..”. Yes, these simulated datasets stem from the LPJ-wsl v2.0 age-class simulations, with simulated fire and prescribed land use change and wood harvest,

Reviewer_2_Technical_Comment_009: l.35 “40-Pg C” -> “40 Pg C”

> Edited accordingly.

Reviewer_2_Technical_Comment_010: l.35-36 A 40 Pg C increase over which time period?

> Over the full simulation period the live biomass carbon in the no_age simulation is greater by ~40 Pg C, as compared to the age-class simulation. In the age-class simulation, there soil carbon is greater by ~33 Pg C and litter carbon greater by ~7 Pg C, as compared to the no_age simulation.
>
> We revised as below. (bold for emphasis, here only).
>
> “..and a finding of a 40 Pg C increase **in biomass turnover when including age dynamics** at global scales..”

Reviewer_2_Technical_Comment_011: l.38 “upper limit” – what do you mean with upper limit? That the forest will not be younger? Please consider rephrasing/explaining. And couldn’t the modelled disturbances (fire, harvest and land-use changes) also be too strong in some grid-cells

leading to forest which is too young (particularly due to the applied old forest first rule; see also general comments)?

We removed the term 'upper limit' throughout. We added clarifying text as below.

We edited the sentence in the abstract to the following, "The LPJ-wsl v2.0 age-module represents another step forward towards understanding the role of demography in global ecosystems."

We added the following text in the Discussion, "In some locations, it is possible that our wood harvest priority rules (harvest oldest age-class first) might lead to simulated stand ages that are younger than observed stand ages if other harvest rules were applied in practice, such as preferentially logging forests of intermediate age with a goal of preserving the oldest forests from harvest. We evaluated the age distribution by continent simulated by LPJ-wsl v2.0 to the Global Forest Age Database (GFAD v1.0, Poulter et al. 2018), which is derived from country-level inventory data (SM Figure 11). The comparison shows that the simulated ages are consistently older than the GFAD dataset."

Note that the GlobFIRM fire module definitively underestimates burned area, see data for fire-model intercomparisons from FireMIP (Hanston et al. 2020); we have updated the Hanston et al. 2017 reference to the 2020 paper throughout. The FireMIP results confirms similar findings about GlobFIRM.

Reviewer_2_Technical_Comment_012: l.41-45 this seems to rather be an enumeration than a sentence and pretty long, could it maybe be taken apart and rephrased?

We edited as suggested.

Reviewer_2_Technical_Comment_013: l.49-52 This sentence seems imprecise to me: From which of the publication exactly do the ~60% total sink stem from? Over which time period? What are the time periods for which Pan et al. 2011b and Pugh et al. 2019a report/estimate the specified sinks, respectively? Is this in combination with changes in environmental forcings?

60% is an approximation as to the role of regrowth in the global land carbon sink. It is correct to suggest that this is not a settle estimate. We edited the text as below to provide greater clarity.

"On global scales, forest age is a considerable factor in global carbon cycling and the total land carbon sink (3.2 ± 0.8 Pg C yr^{-1} for years 2008-2017; Le Quere et al. 2018). Regrowth following disturbance is comprises a large fraction of the land sink based on estimates of the global regrowth flux from country-level forest inventories (Pan et al. 2011a; tropical regrowth sink of 1.6 ± 0.5 Pg C yr^{-1} from 1990 to 2007). A multi-model global regrowth analysis, for which LPJ-wsl v2.0 contributed, estimated a global regrowth sink of 0.3 to 1.1 Pg C yr^{-1} due to demography alone over years 1981-2010 (Pugh et al. 2019a)."

Reviewer_2_Technical_Comment_014: l.50 Pan et al. 2011b not 2011a according to the references?

> Correct. The Pan et al. references have been reordered so that 2011a comes first in the text. The references have been updated throughout.

Reviewer_2_Technical_Comment_015: l.51 really 0.3 to 1.1 PgCyr-1?

> Yes.

Reviewer_2_Technical_Comment_016: l.51 When I understood it correctly than the findings in Pugh et al. 2019a are mainly build on exactly the model being described in this study? In this case I would find the line of argumentation circular, in-transparent and therefore somehow scientifically concerning.

> Yes, a version of the age-module was applied in Pugh et al. 2019a. This is now stated for transparency. The age-module was never fully described or presented elsewhere previously.

Reviewer_2_Technical_Comment_017: l.54 why is fire listed separately of "disturbances"?

> In the line referenced, we removed the term 'disturbances' and simply stated '..land use change and land management, and fire ..'

Reviewer_2_Technical_Comment_018: l.60 but have a look at e.g. Zaehle et al. (2006) or Bellassen et al. (2010)

> Following a similar comment from Reviewer #1, we rephrased as below.
>
> "... Following a call to the science community to improve demographic representation in models (Fisher et al. 2015), there is now a growing list of global models that are capable of simulating global ecosystem demographics (Gitz and Ciais 2003, *Model*: OSCAR; Shevliakova et al. 2009, *Model*: LM3V; Haverd et al. 2014, *Model*: CABLE-POP; Lindeskog et al. 2013, *Model*: LPJ-GUESS; Yue et al. 2018, *Model*: ORCHIDEE MICT; Nabel et al. 2019, *Model*: JSBACH4), although more models need the capability to represent landscape heterogeneity in forest structure and function. ..."

Reviewer_2_Technical_Comment_019: l.64-65: Unfortunately, I cannot find this order in Frolking et al. (2009). In section 3.1 in Frolking et al. (2009) globally disturbed fire area is largest (∼3 × 106 km2 a−1) but only 1 × 105 km2 a−1 in forest – which is equal to that estimated for wind (∼1 × 105 km2 a−1), while global estimates for wood harvest and shifting cultivation are larger–each ∼1–2 × 105 km2 a−1 of forest area.

> We removed the sentence from the text. We were referring to general disturbances over all ecosystems. The reviewer is correct in the Frolking reference.

Reviewer_2_Technical_Comment_018: l.76-79: Please clarify: why forest management here – elsewhere land-use change and land management?

Edited text, changed 'forest management' to 'LUCLM'

Reviewer_2_Technical_Comment_020: l.81 Could you specify which gap exactly? Else maybe omit this phrase?

The sentence was rephrased as below.

"The overall aims of this study are to present new model developments that simulate the time-evolution of age-class distributions in a global ecosystem model and ..."

Reviewer_2_Technical_Comment_021: l.81-83: Note: Several of the studies listed in l.55-57 have demonstrated that a representation of demography influences ecosystem stocks and/or fluxes.

We changed the text to clarify as below. (bold for emphasis, here only).

"...to determine if explicit representation of demography **in this model** influenced ecosystem stocks and fluxes..."

Reviewer_2_Technical_Comment_022: l.85 is there any more recent reference than Sitch et al. 2003 (maybe Poulter et al, 2015; Zhang et al. 2017,2018)? Or maybe rephrase e.g. "a model building/based on the Lund..."?

Sitch et al. 2001 is the main reference for the LPJ model. Bondeau et al. (2007) provide technical details for additional advancements, namely the agriculture module, but we do not use the agriculture module in this paper. The development history is described in Section "2.1.1 LPJ History"

Reviewer_2_Technical_Comment_023: l.110 are?

Edited accordingly.

Reviewer_2_Technical_Comment_024: l.115 before and elsewhere in the text I understood that fire is also implemented as a stand replacing disturbance/ burned fraction moves to youngest age-class?

For clarity, we changed two mentions of 'stand replacement' or 'stand-clearing' in the text as below.

".. Although pest and pathogens, namely bark beetle infestations, affected a much larger area (up to 6% of total forested area in U.S.) than both logging and fire, their effects do not always cause ~~stand replacement~~ immediate tree mortality. .."

".. Not all trees are killed-off when a ~~stand-clearing~~ disturbance occurs in LPJ. ..."

Reviewer_2_Technical_Comment_025: l.130 "unequalbin setup is applied to explore model dynamics at the level of a single grid-cell;" according to Table2 its not a single grid-cell but region, which is also suggested by e.g. Fig.4.

We edited the text as below.

".. The *10yr-equal*bin age setup is used for all simulations including the global simulation, whereas the *unequalbin* setup is used for regional and single grid cell simulations; simulation details in next section. ..."

Reviewer_2_Technical_Comment_026: l.127-131: I would appreciate a bit more information on and explanation of the choices that drove the separation in age-classes. Particularly, why is the cut off with 151years in the 10-yr equal bins and why is it with 101years in the unequal bins? Why is the age range of the pre-last class (code 11) in the 10-yr equal bin larger – making it an "unequal bin", too. Maybe also the motivation for the 2, 5 and 25 year ranges as well as the switches between these ranges could shortly be outlined? If this resulted e.g. from preliminary tests, the experiences of the authors could maybe be instructive to the readers.

The binning was chosen to align with U.S. forest inventory data and we wanted greater resolution in the age-classes between 1-100. The unequal-bin setup was primarily implemented to evaluate issues with the equal-bin setup. We did not explore other binning methods as we were satisfied that the equal-bin setup was sufficient. We added text to clarify why we use the 10-year equal bin setup for global simulations as below.

".. The use of equal or unequal age class setups is more than just for reporting purposes. Resources available to plants (space, light, soil water) differ between age-classes but not within age-classes, and we limit the model to represent a total of 12 ageclasses only. Also, there exists a greater range of forest ages at global scales and the equal age-class setup allows us to independently model resource dynamics for more of the terrestrial surface. If we had chosen the unequal-bin setup for global simulations, we would be independently modeling processes only for the youngest age-classes and we would lose capacity to independently model processes at intermediate and older age-classes. .."

Reviewer_2_Technical_Comment_027: l.146 . . . number "of" simulated . . .

Edited accordingly.

Reviewer_2_Technical_Comment_028: l.161 I would recommend to introduce a j on the w to indicate that the age-classes (can) have different widths EQ4 and l.173 personally I find f0,0 an unlucky choice and would prefer an extra term, such as fdis or the like EQ5 and l.179 why a capital F in F(t)_w,j-1, isn't this just one entry?

We changed f0,0 to f_{new} clarity. Yes we reference a single element; we changed the text accordingly to $f_{w,j-1}^{(t)}$

We revised Equation 2 to show that the sum of fractional areas for all patches in a grid cell is defined by the sum of fractional areas for all age classes and age widths.

Reviewer_2_Technical_Comment_029: l.192-199: is this an enumeration? If so, could it maybe be separated with newlines? Else I would appreciate complete sentences.

Prior text in reference "Within-class Fractional Transitions: For every simulation year, the position of each element (fx) in the VTFT vector is incremented by the representative time of each element (x), which is simply 1. No changes occur to the state variables of the age-class during within-class transitions. Between-class Fractional Transitions: Upon incre195 menting the position of each element, if the value at (fw) is non-zero, then the corresponding fractional area fw, defined as the outgoing fraction, is used in an area-weighted average between the state variables of a1 fw and the next oldest age-class a2 F_total. Lastly, upon incrementing element position, if all elements < f1 ... fw > in the VTFT vector of the preceding age-class, in this example (a1), are zeros, then the age-class is simply deleted from computational memory."

Text above changed as below, with

"The following is a description for within-class and between-class transitions. *Within-class Fractional Transitions*: For every simulation year, the position of each element (f_x) in the VTFT vector is incremented by the representative time of each element (x), which is simply 1. No changes occur to the state variables of the age-class during within-class transitions. *Between-class Fractional Transitions*: Upon incrementing the position of each element in the VTFT vector, if the value at f_w is non-zero then the corresponding fractional area (f_w), defined as the outgoing fraction, is used in an area-weighted average between the state variables of a_1 *fw* and the next oldest age-class a_2 *F_total*. Upon incrementing element position, if all elements in the VTFT vector of the preceding age-class are zeros then the age-class is simply deleted from computational memory."

Reviewer_2_Technical_Comment_030: l.202 to which age-widths does this refer to, those from the unequal setup or both setups? Is there a specific section of the manuscript where "it is demonstrated" or is this a more general statement as "in this study"?

The text was edited for clarity, as below.

"Two hypothetical scenarios are provided in Figure 2 that demonstrate age-class transitions using the VTFT procedure when there is a young age-class created, and when there are fractional age-class transitions between age-classes. With VTFT, any number of age-classes and age-widths can be modeled, but it is demonstrated in this study that ..."

Reviewer_2_Technical_Comment_031: l.206 and 220 "merged with a youngest" -> the? Or can there be several youngest?

'youngest' is use in the singular, I'm not sure there is a plural interpretation to the word. For plural, one might say the 'young' ageclasses.

Reviewer_2_Technical_Comment_032: l.213 I do not understand this, why can't the not burning fraction stay in the current age-class/patch and only the burned fraction move to the youngest age-class?

Only the fraction that burns gets moved to the youngest age-class. The fraction that does not burn stays in the current age-class. The text (L 213) refers to the PFT population that does not burn completely and kill-off all the trees. The simulated burned fraction may have surviving trees.

Reviewer_2_Technical_Comment_033: l.215 This is the first time primary and secondary forest are mentioned. Also, the term tile has only been mentioned one time before ("Age-classes are represented as subtiles within a grid-cell"). Maybe it would help to already introduce these aspects in 2.1.1?

Removed the sentences below.

~~Fire can occur in both the primary forest and secondary forest tiles; the classification of primary versus secondary forests is determined by the land use driver dataset~~

We have replaced most instances of 'patch' and 'tile' with simply 'age-class' throughout.

Reviewer_2_Technical_Comment_034: l.217 Does managed land refers to crop/pasture here (i.e. not forest management)?

Yes. Text edited for clarity as below.

"Age-classes get created when managed land (i.e., crop/pasture) is abandoned..."

Reviewer_2_Technical_Comment_035: l.225-226 mix of singular and plural?

'give' changed to 'gives'.

Reviewer_2_Technical_Comment_036: l.224 I assume this is not a only "if the". Consider rephrasing such that it gets apparent that net zero land-use change is just one example?

Not sure I understand the Reviewer's comment, but we edited the text as below.

"In the LUCLM module, gross transitions between land uses are simulated (Pongratz et al. 2014, Stocker et al. 2014), such that if the fraction of abandoned land equals the fraction of land deforested in the same year (net zero land use change) then the fluxes from the gross transitions are tracked independently and give an overall more accurate

accounting (and higher magnitude) of emissions from LUC than if we only tracked net transitions. ...”

Reviewer_2_Technical_Comment_037: l.229, l.263, l.527, l.602 consider updating to Hurtt et al. 2020 l.228 lost “and”?

The citations were updated to Hurtt et al. 2020 as suggested.

Reviewer_2_Technical_Comment_038: l.229: I do not understand what you mean with modifications 1a (and 2a) seem not to be modified with respect to LUH2?

The text was updated for clarity, as below.

“... but with the following modifications so that the LUHv2 data can be used in LPJ-wsl v2.0: ...”

Reviewer_2_Technical_Comment_039: l.233: LUH2 offers a separation of harvest to mature and young forests. Consider shortly stating why this separation is not used in LPJ-wsl 2.0?

LUHv2 does not provide distinction of stand age at finer granularity other than ‘young’ and ‘mature’.

Reviewer_2_Technical_Comment_040: l.233-237: But wouldn’t e.g. shifting cultivation rather make use of younger forests?

It is possible, yes.

Reviewer_2_Technical_Comment_041: l.237-238: LUH2 offers both, harvested area and harvested biomass. Here it is stated: “until two conditions are met” and in the next sentence: “until a prescribed harvest mass or harvested area is met”. This requires clarification when which of these criteria is applied.

(1b) refers to land-use change, and land-use change is prescribed by an areal fraction. We clarified as follows, “(1b) For simplicity, deforestation (i.e., land-use change) ..”

(2b) refers to wood harvest, and wood harvest is prescribed by an areal fraction and the biomass harvested on that fraction. We clarified as follows, “(2b) wood harvest (i.e., biomass harvest) also occurs in the ranking of oldest to youngest age-class ...”

Reviewer_2_Technical_Comment_042: l.244-245: I wonder if this would really be the case, I would assume that the ranking from old to younger age-classes decouples deforestation and abandonment?

The text (L 244-245) refers to a computational issue involved when modeling gross transitions. In a single year, crop/pasture can be abandoned (converted to secondary

forest) and forest can be converted to crop/pasture. The order in which these processes are simulated will introduce a bias, or more aptly a model artefact.

Reviewer_2_Technical_Comment_043: l.240-. . . Is this new in LPJ-wsl 2.0 or is this as it has been done already before? Noticing Earles et al (2012)and McGuire et al. (2001) in 249/251 I wondered if the authors could also give the reference for the % ratios in l.240-247?

This is new in LPJ-wsl v2.0. The reference for Earles et al. (2020) is with regard to the concept of delayed emissions. The 40:60 rations are from McGuire et al. (2001), as referenced.

Reviewer_2_Technical_Comment_044: l.249-251: could you clarify which numbers are from Earles?

See above.

Reviewer_2_Technical_Comment_045: l.251-256: "product pool" is used twice here – with different meanings?

We think they have the same meanings. In the literal sense, 'product pool' means 'pooling products'. Another way of interpreting it would be 'carbon stock in [wood] products', or 'storage container for carbon products'.

Reviewer_2_Technical_Comment_046: l.253 "dataset described further in Sect 2.3.3" – I cannot find such a description there?

We removed the text " ~~dataset described further in Sect. 2.3.~~"

Reviewer_2_Technical_Comment_047: l.265 "managed lands" = agricultural managed lands (since forests can also be managed?)

'Managed land' has different meanings in reference to LUHv2. We specify in the following sentence in the text that "In LPJ-wsl v2.0, managed lands (i.e., crop/pasture) are treated as grasslands with no irrigation, no fire, and tree PFTs were not allowed to establish."

Reviewer_2_Technical_Comment_048: l.271 "accessed . . ." consider moving to references.

We decided to kept as is.

Reviewer_2_Technical_Comment_049: l.272 "fuzzed" is this relevant for this study?

Yes.

Reviewer_2_Technical_Comment_050: l.275 Refer to SM2 here.

Edited accordingly.

Reviewer_2_Technical_Comment_051: l.275 “model-observation comparisons” – isn’t the model resolution anyway 0.5◦ in the compared simulations?

Yes. By aggregating the plot data to larger domains (USFS Divisions) we intended to reduce the potential influence of differing climate, soils, and location between simulated data and the observations. Such aggregation has been done before for similar reasons, see Purves et al. 2008 PNAS and its supplementary materials.

Reviewer_2_Technical_Comment_052: l.310 and regrowth?

We left as is.

Reviewer_2_Technical_Comment_053: l.318 info does not match Table 2.

Yes it does. Section 2.3.4 refers to the single-cell simulation, not the Regional simulation.

Reviewer_2_Technical_Comment_054: l.320 it is unclear to me which of the deforestation rules from Section 2.2.2. also applies for the Snoage_event simulation, could you please give a bit more detail?

We rephrased as below for clarity.

“Treatment of deforestation byproducts (i.e., carbon in dead wood left on-site) were the same in both simulations.”

Reviewer_2_Technical_Comment_055: l.322 NBP so far not introduced (NPP and Rh only in the abstract).

We clarified the text as follows, “Net Biome Production (NBP, defined as NBP = NEP – LUC_flux)”

Reviewer_2_Technical_Comment_056: l.330 Table lists 4 objectives/questions.

We edited the text accordingly, as below.

“The third and fourth objectives used data from $\mathbf{S_{age}}$ to determine where the effect of demography was greatest and to identify the relative influence of demography versus climate on simulated fluxes (NEP, NPP, and Rh).”

Reviewer_2_Technical_Comment_057: l.336 Maybe already add here for clarification that Sage = SFireLU.

Edited accordingly.

Reviewer_2_Technical_Comment_058: l.339 “all three simulations” presumably refers to SFire, SLU and SFireLU? But what about Snoage? What was the spin-up procedure for this simulation?

We edited the text to clarify as follows, “For all global simulations..”

Reviewer_2_Technical_Comment_059: l.339 does the first spin-up also has “land use values” or does it assumes only natural vegetation?

We clarified as follows, “a spinup simulation...and no land use or wood harvest...”

Reviewer_2_Technical_Comment_060: l.341 Could you please specify what you mean by ‘natural conditions’ – fire?

Some of the simulations had fire turned-off. We removed the text in reference (‘natural conditions’) and the sentence now reads as follows, “..spinup ensured that age distributions and state variables were in dynamic equilibrium (i.e., no trend).”

Reviewer_2_Technical_Comment_061: l.342 please clarify “land use values” does that mean managed agricultural land distribution? What about harvest?

We restated for clarity as below, “..to initialize land use fractions of crop/pasture to year 1860..”. Wood harvest was not simulated during spinup procedures.

Reviewer_2_Technical_Comment_062: l.342 please clarify: was the second spin-up procedure subsequently or alternatively for different simulations? Do all four simulations start from the same values in 1860?

Only simulations that used land use had a second spinup procedure. We edited the text to clarify as follows, “For simulations with land use, a second ‘land-use-spinup’..”

Reviewer_2_Technical_Comment_063: l.356-359 I found this sentence a bit difficult to read since the “By contrast, fire . . .” seems to refer to the “Trends in LULCM are . . . prescribed” – please clarify by e.g. rephrasing.

Yes, “By contrast, fire is a fully simulated process . . .” does refer to the “Trends in LULCM are . . . prescribed”. We rephrased for clarity as below.

“Trends in age distributions due to LUCLM are not prescribed by inputs per se; instead, the age module is a necessary model structure that allows full realization of the effect of forcing data on age distributions. Trends in age distribution due to Fire, which is a simulated process as opposed to prescribed, result from climate and fuel load feedbacks on fire simulation.”

Reviewer_2_Technical_Comment_064: EQ6 I wonder if the last factor should be written as a sum with age classes as index?

The equation was edited as below, with the 'age' subscript in the ageclass term removed.

$$flux_{i,yr} = B1_i \times \text{total\_precipitation}_{i,yr} + B2_i \times \text{mean\_temperature}_{i,yr} + B3_{i,ageclass} \times ageclass_{i,yr}$$

In any manner the last term in Eq #6 is correct, it is not a sum. In the equation, the indices refer to how the terms vary. In the case of the last term in the equation, $B3_{i,age} \times ageclass_{i,yr}$, the beta coefficient (B3) vary as a function of grid cell (i) and ageclass code (*ageclass*). The ageclass code can vary as a function of grid cell (i), year (yr).

Reviewer_2_Technical_Comment_065: l.393 "age-structure patterns" – maybe "patterns of tree density and height per age"?

Yes, that was the meaning. We edited the text accordingly.

Reviewer_2_Technical_Comment_066: l.397 what does stand refers to –patch?

Yes. 'patch' changed to 'age-class' for consistency.

Reviewer_2_Technical_Comment_067: l.404 I do not understand this part of the sentence: "data be taken on every species; although species-level data are available".

The paragraph has been edited for clarity of meaning. The updated text is below.

"FIA data had greater variability among age-classes, regardless of Division. FIA data are not aggregated by PFT, instead they are species-level data. Changes in species composition over time do occur and it can add to the observed variability among age-classes in tree density and tree height. LPJ-wsl v2.0 includes a limited set of PFTs, which most likely limits the model's capacity to represent similar levels of variation in tree density and tree height. It is beyond the scope of this study to disentangle these patterns further, but greater agreement between observed and simulated patterns of forest structure might be acheived by including additional plant functional types that are representative of tree species for a given Division."

Reviewer_2_Technical_Comment_068: l.417 These survivor trees make me think if a classification as "time since disturbance" would make more sense than a classification as age-classes?

We maintain that the 'age-class' terminology is the correct terminology to use. Stand age ('age-class') is typically used to explain consistent and predictable patterns of ecosystem function and forest structure with stand age, survivor trees notwithstanding.

In the simulations, survivor trees represent a very small fraction of the PFT population. In Figure 3, the survivor trees are evident in lowest age-class of the tree height plot for the unequal-bin simulation setup. The corresponding tree density in the lowest age-class is

very low. Survivor trees are likely also present in the equal-bin setup, but the patterns for tree density and tree height are not affected.

Reviewer_2_Technical_Comment_069: l.417 LPJ? LPJ-wsl 1.0? LPJ-wsl2.0? All of them?

For consistency, we changed all references to the current model to LPJ-wsl v2.0.

Reviewer_2_Technical_Comment_070: l.442 all "three" U.S. States.

No, not 'three' states, but *all* states. The text in reference (L 442) refers the consistent and predictable pattern of NEP as a function of age. We estimated the exponent using data pooled over all states, and separately using data from each state. The exponent value was consistent.

Reviewer_2_Technical_Comment_071: l.452 Figure 5?

Yes. The text was edited accordingly.

Reviewer_2_Technical_Comment_072: l.457 missing t in event.

We cannot find the typographic error in question.

Reviewer_2_Technical_Comment_073: l.457 LPJ? LPJ-wsl 1.0? LPJ-wsl2.0? All of them?

We changed all references to the current model to 'LPJ-wsl v2.0'.

Reviewer_2_Technical_Comment_074: l.484 "(?)"?

We edited the sentenced reference as follows, "The question still remains – should there be an expectation for greater differences in NEE?"

Reviewer_2_Technical_Comment_075: l.487 Snosge -> Snoage

Edited accordingly.

Reviewer_2_Technical_Comment_076: l.498 LPJ? LPJ-wsl 1.0? LPJ-wsl2.0? All of them?

We changed all references to the current model to 'LPJ-wsl v2.0'.

Reviewer_2_Technical_Comment_077: l.508-509 the 23 years are not directly evident from Table 3, nor is it the decrease in zonal ecosystem age, could you help your readers specifying which of the values in Table 3 show these? This also holds for the rest of the paragraph; maybe consider extending Table 3 or adding another table showing integrated values?

We rewrote the paragraph to clarify as below. Originally, we arrived at the '23 year' difference in ages by expanding the statistical model to estimate the ecosystem age in year 1 (simulation year 1860). As an example, the age in year 1 for Boreal latitudes is given by `(141.7−(0.0098∗1)) = 141.6` , the age in year 1 for Temperate latitudes is given by `(118.5−(0.0525∗1)) = 118.4`, and the difference is given by `141.6 − 118.4 = 23.2 [years]`. To arrive at the age estimated in the simulation year 2016, the year index is 157, so the age would be given by `(141.7−(0.0098∗157)) = 140.1614`.

"Ecosystem age by zonal band was oldest at boreal latitudes, followed by temperature latitudes, and youngest in tropical latitudes, which was primarily the results of frequent fires in simulated grassland ecosystems. The primary driver of zonal age distributions was Fire (Figure 8). According to results from the statistical model (Table 3), the average age difference due to fire among zonal bands in 1860 was 23 years between Boreal (older) and Temperature (younger) latitudes, and it was 32 years between temperature (older) and tropical (younger) latitudes. The difference in ecosystem age among zonal bands increased to 60 years in simulation year 2016 between boreal and temperate latitudes, while the difference in ages between temperature and tropical latitudes remained similar (31 yr age difference). There was a statistically significant decrease in zonal ecosystem age over time due to fire (Table 3)"

Reviewer_2_Technical_Comment_078: l.519 grammar issue?

Not sure we understand the Reviewer's concern with the sentence in question.

Reviewer_2_Technical_Comment_079: l.529 also here sum over B3age?

No, there is no summation in the statistical model.

Reviewer_2_Technical_Comment_080: l.530 simulated NPP and Rh.

Yes, we added 'simulated' to clarify we mean the simulated NPP and Rh.

Reviewer_2_Technical_Comment_081: l.533 consider to delete "slightly"!

Edited accordingly. The difference is more than 'slight', thanks.

Reviewer_2_Technical_Comment_082: l.587 LPJ? LPJ-wsl 1.0? LPJ-wsl2.0? All of them?

We changed all references to the current model to 'LPJ-wsl v2.0'.

Reviewer_2_Technical_Comment_083: l.593: again I would recommend clarifying "upper limit" and again I am not sure if this is correct, due to the oldest age-classes first principle for harvesting and deforestation (l. 233-239).

Per previous response, see below.

We removed the term ‘upper limit’ throughout. We added clarifying text as below.

“In some locations, it is possible that our wood harvest priority rules (harvest oldest age-class first) might lead to simulated stand ages that are younger than observed stand ages if other harvest rules were applied in practice, such as preferentially logging forests of intermediate age with a goal of preserving the oldest forests from harvest. We evaluated the age distribution by continent simulated by LPJ-wsl v2.0 to the Global Forest Age Database (GFAD v1.0, Poulter et al. 2018), which is derived from country-level inventory data (SM Figure 11). The comparison shows that the simulated ages are consistently older than the GFAD dataset.”

Reviewer_2_Technical_Comment_084: l.606 same here with the underestimation – given the oldest age-classes first principle I am not fully persuaded that it underestimation is granted.

Per previous response, see below.

We added text to clarify that we think our simulations represent the upper bound of age distributions, where ‘bound’ is meant to convey a range of values [lower, upper] of expectation. We added clarifying text as below.

“In some locations, it is possible that our wood harvest priority rules (harvest oldest age-class first) might lead to simulated stand ages that are younger than observed stand ages if other harvest rules were applied in practice, such as preferentially logging forests of intermediate age with a goal of preserving the oldest forests from harvest. We evaluated the age distribution by continent simulated by LPJ-wsl v2.0 to the Global Forest Age Database (GFAD v1.0, Poulter et al. 2018), which is derived from country-level inventory data; we have added this comparison as a figure to the Supplement. The comparison shows that the simulated ages are consistently older than the GFAD dataset.”

The GlobFIRM fire module underestimates burned area, see data for fire-model intercomparisons from FireMIP (Hanston et al. 2020); we have updated the Hanston et al. 2017 reference to the 2020 paper throughout. The FireMIP results confirms similar findings about GlobFIRM.

Reviewer_2_Technical_Comment_085: l.622 But isn’t this model dependent? Maybe consider rephrasing, e.g. “suggesting that uncertainty in carbon residence time could potentially be reduced” or the like

Agreed, it can be model dependent. We added the clarifying text ‘..could potentially be reduced..’ as suggested.

Reviewer_2_Technical_Comment_086: l.624-627: I do not agree that this is “the current state of knowledge”, nor that “existing models that estimate the global land-use flux... do not include age dynamics”. For the former and the latter please e.g. refer to findings of Yue et al. 2018, in addition, for the latter, the authors might have a look into other studies conducted with some of the models listed around l.55-57.

We respectfully disagree. Yue et al. 2018 report on a single grid cell (0.5 degree) idealized simulation and it does not represent consensus on the state of knowledge. We refer to global emissions from gross land use change being greater than net land use change based on Arneth et al. 2017.

Reviewer_2_Technical_Comment_087: l.631: consider adapting the subsection header since this subsection seems to be more about precipitation than demographic effects?

We added some text to the section in reference so that the section title reflects the content of the paragraphs therein.

Reviewer_2_Technical_Comment_088: l.662: is this only the case if using the unequal age-class setup?

No, NEP is greatest in the youngest age-class, regardless of the simulation setup.

Reviewer_2_Technical_Comment_089: l.664: LPJ? LPJ-wsl 1.0? Both?

We changed all references to the current model to 'LPJ-wsl v2.0'.

Reviewer_2_Technical_Comment_090: l.671: I assume this is the case in several of the models listed around l.55-57.

We aren't sure, one way or another, if this applies to other models. We know this does apply to LPJ-wsl v2.0.

Reviewer_2_Technical_Comment_091: l.675: consider adding "on the same machine" (if this is correct).

Not correct. LPJ-wsl v2.0 can be run distributed, in parallel, on multiple compute 'machines'.

Reviewer_2_Technical_Comment_092: l.680: The first 2-3 sentences seem to be incomplete?

The section was re-written and begins with the statements as below.

"In order of priority for improvement of the age-module: 1) improve age-class growth rates to align with observations, 2) improve representation of disturbances, 3) improve representation of early- and late-successional plant species and add vertical structural complexity such as understory/overstory canopy. Below, we provide suggestions and examples from the literature as how these improvements might be accomplished."

Reviewer_2_Technical_Comment_093: l.700: To my understanding JSBACH4 does not represent much vertical heterogeneity. You might want to have a look into e.g. ORCHIDEE-

CAN (Naudts et al. 2015) or in individual based models (in addition to ED), e.g. LPJ-Guess (Bayer et al., 2017).

> The phrasing 'much' is relative. Relative to LPJ-wsl v2.0, which has a single layer canopy, JSBACH4 provides a good example for future developments.

Reviewer_2_Technical_Comment_094: Table 1: LPJ-wsl v2.0?!

> As above, we changed all references to the current model to 'LPJ-wsl v2.0'.

Reviewer_2_Technical_Comment_095: Table 2: * single-cell: included processes might not match the description in 2.3.4, l.318. * global: Initially I tried to associate each of the four questions with one of the simulations, due to the visual structuring of the rows of the table. Maybe merge cells and number questions?

> We choose to leave as is. The questions are meant to be interpreted as 'objectives and questions', as in general questions we wished to answer using the global simulation data. There are other questions we address in the text for each simulation. The column was not meant to be a full enumeration of all research questions and associated findings we address within the main text.

Reviewer_2_Technical_Comment_096: Figure 1: please explain what you mean with "land use" in this context

> In the context of LPJ-wsl v2.0, 'Land Use' refers to crop/pasture.

Reviewer_2_Technical_Comment_097: Figure 3: explain MI,MN & WI again.

> We clarified as follows, "..for U.S. States of MI, MN, MN.."

Reviewer_2_Technical_Comment_098: Figure 5: consider increasing visibility by changing the y-axis of the first panel (max of -5/-6 kgCm-2).

> We leave as is. We think the important content of the panel is sufficiently displayed while leaving room for the legend.

Reviewer_2_Technical_Comment_099: Figure 6: *consider adding simulation names (Sage and Snoage if I understood it correctly). * could you show the simulation starting from the spin-up, i.e. starting 1860? Is the difference between the simulations due to the spin-up or evolving in the course of the simulation?

> The simulations have the same prescribed drivers (inputs). The differences between simulations are observed after spinup, yes.

Reviewer_2_Technical_Comment_100: Figure 7: * consider using (a) and (b) instead of left and right. * “LPJ-wsl simulations” consider adding simulation name from Table 2.

We kept ‘left’ and ‘right’.

Reviewer_2_Technical_Comment_101: Figure 9: Since SFireLU is more complete than SFire, consider using the solid line for this more complete set-up?

We left as is. We annotate each line in the plot with the associated simulation.

Reviewer_2_Technical_Comment_102: Figure 10: * consider using the same y-axis for better comparability (same SMFig1). *

We left as is.

Reviewer_2_Technical_Comment_103: l.988 model is can -> model can

Edited accordingly.

Reviewer_2_Technical_Comment_104: Figure 11: * l.993 “black is zero”? On the colour map it is yellow? * red (-0.3) and pink (0.7) are difficult to distinguish, maybe consider a change in the colour map.

Black is zero, black is not yellow, it is stated as such in the legend. We provide the actual datasets for individual inspection, so we decided to leave the color scheme as is.

Reviewer_2_Technical_Comment_105: Figure 12: consider labelling panels (a)-(d) instead of using top row, bottom row, top left and bottom left.

We prefer the top/bottom, left/right referencing. We don’t reference every panel in the plot in the text, otherwise this would be a good suggestion.