Response to Reviewer #1

Reviewer_1_ Main_Comment_001: Disturbance history and disturbance regime are important drivers of terrestrial biosphere dynamics and ecosystem function, but they are rarely represented in dynamic global vegetation models. Here Calle and Poulter describe their age-class implementation in the LPJ model (LPJ-wsl v2.0), and present a series of simulations seeking to highlight the effects of disturbance history on vegetation structure and the carbon cycle, as well as the global patterns of ecosystem age when accounting for fire and land cover and land use disturbances. This work provides an important model development and can become an important contribution to the modelling community, once some issues, which I describe below, are addressed by the authors.

The current model description provides an overview of the age structure in LPJ-wsl and includes some examples on how this module works (Figures 1 and 2). However, some mechanisms are not sufficiently described and deserve attention, especially in a journal like GMD. For example, in section 2.2.1, I could not tell how each within ageclass element ($f_{i,j}$) is represented in the model: are they treated as "independent" components (i.e., available soil water and light computed independently for each within age-class element), or do all the elements in the same age class share the resources?

> Agreed, this could be clearer. The hierarchical structure of the model is described on L120. All ageclasses share the same gridcell inputs (climate, co2, radiation). The state variables of plant available soil water and light can differ among ageclasses, which is mainly controlled by plant water demand and plant cover, respectively.
>
> The *within* ageclass elements are not independent and every *within* ageclass element has the same exact state variables, including the same soil water and light. The *within* ageclass elements are simply a vector representation of areas for each age-unit in the ageclass. As such, we only simulate processes at the ageclass level, and the *within* ageclass elements are a simple method for a 'smooth' transition between ageclasses (i.e., no big jumps in state variables when ageclasses transition). In theory, we can simulate processes independently for each within ageclass element, but this is not practical or necessary.

Also, how do the age-width transitions work in the case of unequal age classes, considering that the age class transitions occur once a year? Does that mean that young age classes have fewer elements, or are multiple elements allowed to transition to another age class at the annual time step? These are mostly points for clarification and should be straightforward to address in a revised version.

> Correct, ageclass transitions occur only once per year. In the unequal-bin setup, young age-classes have fewer elements. Each *within* ageclass element represents the areal fraction of a single age-unit, for either setup (equal-bin and unequal-bin ageclasses). Every year, all elements increment, but each element can only increment its position once per year (rate of change is 1). Per response above, the main benefit for using unequal-bin ageclasses is to independently simulate processes and track state variables separately.

We added the text below Section 2.2.1 "An age-based model of ecosystems – sub-grid-cell patch dynamics. (bold is for emphasis, here only).

... The age-class module has a fixed number of age-classes that can be represented in a grid cell, but all age-classes are not always represented. **Age-classes are classified into 12 age-classes (patches) in fixed age-width bins, defined as the *unequalbin* or the *10yr-equalbin* age-width setup (Table 1). Each ageclass contains *within* ageclass elements, which are simply a vector representation of areas for each age-unit in the ageclass. 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 age-widths of the age-classes in the *10yr-equalbin* setup correspond to common age-widths of classes used in forest inventories. The *10yr-equal*bin age setup is used for all global simulations, whereas the *unequalbin* setup is applied to explore model dynamics at the level of a single grid-cell; simulation details in next section.

The authors compare the effect of some model settings (e.g., enabling vs. disabling the age structure module), but no benchmarking is provided other than the comparison of the predicted forest structures with FIA plots. Consequently, several processes were not truly evaluated against observations or at least reported values in the literature. For example, when the authors compare the simulations with and without age-class dynamics (Figures 5, 6 and text referring to them), it is implied that the age-structure simulations are more reasonable, but the authors do not provide any reference to observations. Although the simulations are idealized, some values from literature could at least indicate whether the time scale for recovery is at least in the right order of magnitude at different biomes.

There is value in improving modeled forest structure. The comparisons to FIA plot data are intended to provide confidence in the model's capacity to reproduce forest structural properties – a form of benchmarking. We provide a new comparison of the age distribution by continent simulated by LPJ-wsl v2.0 and compared 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.

This work is not intended to be a benchmarking/optimization paper, although we intend to do this in the future. Benchmarking, optimizing model parameters, identifying and improving model processes is no small task. Throughout the Discussion sections we use phrasing throughout that accounts for our uncertainty in our simulation results. We make this clear in the first section of the Discussion and we add comments to clarify our uncertain position, as below.

"(4.1) Distribution of Ecosystem Age on Earth

... Our model developments are not optimized to match observations, although we are working toward this end. Future goals are to assimilate stand-age related data, such as remotely-sense canopy data and stand index growth curves, to align model processes with observations. ..."

"(4.2 Age Dynamics Increase Turnover) ... That turnover increases when explicitly simulating ageclasses is a natural expectation, but the magnitude of the simulated turnover between carbon pools is less certain until detailed benchmarking is conducted. ..."

Finally, the fire disturbance is presented as the critical determinant of forest age distribution, but no assessment of the fire module is provided. I understand and agree with the authors that fire datasets such as GFED will include fire types currently not represented in LPJ-wsl and a comparison of carbon emissions is not possible due to the risk of double counting, but they could be still useful for verifying whether spatial distribution and the inter-annual variability of fire disturbance predicted by LPJ-wsl is reasonable or not.

The fire module was left unchanged from prior versions. The Glob-FIRM fire module has been previously evaluated in great detail by Thonicke et al. 2001, and Hanston et al. 2016. There are better efforts at answering the utility or realistic representation of simulated fire dynamics than we can do justice in this paper. This was the major aim of the Fire Model Intercomparison Project (FireMIP). We make it clear that the fire module needs improvement, it underestimates burned area, and that the resultant effect is older ecosystem ages.

In the second paragraph of the Discussion, we added clarifying remarks as below. (bold for emphasis, here only)

"...Furthermore, the fire module has been well evaluated at global scale (Thonicke et al. 2001) but **it needs improvement because it is overly simplistic and underestimates global burned area** (Hantson et al. 2020), so it is more likely that effects of fire are much greater than simulated in this study. It is clear then that this study underestimates disturbances rather than overestimates them, and as such, these simulations overestimate ecosystem age. But again, additional disturbances would only lead to younger age-classes, enhancing the role of age dynamics in regional and global carbon cycles."

Specific Comments

Reviewer_1_Specific_Comment_001: L58. Re-write this sentence, so it is clear that some models do account for demographic effects, including a few that were cited in the previous sentence.

The text was changed accordingly, as below, to clarify that some models already account for demographic effects.

"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."

Note that few models that simulate the global terrestrial surface account for demography. CABLE, LPJ-GUESS and now JSBACH are the few exceptions that now include sub-grid-cell heterogeneity in ecosystem demography. ED2 does have demographic capabilities, as do many other regional and landscape-scale simulation models, but such models have not been run globally. The lack of global simulations demographic models is primarily due to the computational burden of ageclass representation, which we overcome with our methodological approach.

Reviewer_1_Specific_Comment_002: L94. The authors mention permafrost and wetland methane but these features are not described anywhere. Considering that these are features in the new code, shouldn't they be described somewhere?

The LPJ-wsl v2.0 model is written as a fully modular program. Compiler flags are used to turn on/off modules. In this paper, we did not use the wetland methane or permafrost compiler flags.

Reviewer_1_Specific_Comment_003: L132. This is a good and clear explanation, but I wonder if the authors could also highlight the consequences of adding age-classes to the representation of the microenvironments in LPJ-wsl (light, water and perhaps nutrient availability). Also, was there any reason why natural disturbances (e.g., tree fall) cannot create new age classes?

The ageclass module doesn't model microenvironments per se, rather it is intended to represent landscape heterogeneity, but the remark is well taken and it is a good point. The ageclass or patch size is at minimum ~2.5 km2, with a maximum of ~50 km2 (0.5 degree grid cell). Resource availability (space, light, water, *no* nutrients) is implicitly modeled as a function of a mean-individual 'big-leaf' plant functional type (PFT), with each PFT having properties of stem density, fractional plant cover, tree height, and other attributes that govern water demand and space filling properties.

Other disturbances such as tree fall can create new age classes, yes. Our model only includes the disturbances of fire and land use and land management, but other disturbances can certainly be added. The main text has similar phrasing in the Discussion section 4.3 Opportunities for Improving Modelled Age-dynamics."

"..There a number of opportunities for refining the age-module. Incorporating additional disturbances within the model, which will help simulate age distributions more consistent with inventory (Pan et al. 2011a) and satellite (Pugh et 680 al. 2019b) data and contribute to more scientifically relevant questions. Modeled disturbances need not be complex to explore their effects on age distributions, they only need to reset a fractional area to the youngest age-class. ..."

Reviewer_1_Specific_Comment_004: L140–155. This is not entirely accurate. In some cohort-based models, a patch represents a collection of gaps with similar forest structure. In such models, fusing patches that have similar structure simply means that the structures of patch A and patch B are sufficiently similar so that the merged patch can represent all gaps in A and B (and thus representative of a larger area). At least for ED2, the patch fusion is not determined by one state variable as implied in the text, but by the vertical LAI profile (Fisher et al. 2018).

> In ED2, the vertical LAI profile can still be considered a state variable of the patch, even if it is emergent from the underlying PFT cohorts. The point we make is three-fold, (1) some models do not have fixed patch size (LPJ-GUESS has a fixed patch size and patches do not merge); (2) models that have variable patch size require merging similar patches otherwise the patches could be created every year and computation will slow to a crawl. Merging is a computational solution to patch creation. (3) merging patches based on a limited set of state variables, or even a single state variable, is an arbitrary decision along a single axis of similarity between patches. We clarify as below in Section "2.2.1 An age-based model of ecosystems – sub-grid-cell patch dynamics":
>
> "We also employ merge age-classes (patches), but we do not employ merging rules along arbitrary axes of similarity. We fix the number of age-classes *a priori*, similar to LPJ-GUESS in that there is a maximum number of age-classes. Instead of forced merging to reduce computational burden (as in ED2), a fraction of the age-class always transitions to an older state, and a fractional area can transition and merge with the next oldest age-class."

Reviewer_1_Specific_Comment_005: Section 2.2.2. I understand that the fire model has been previously described, but more detail would help here, as fires are critical for the results shown later in the paper. Instead of describing the model qualitatively, the authors could provide the basic equations and also a table with the PFT-specific fire resistances (SI text and table would be fine).

> The fire module is described in greater detail in other papers (Thonicke et al. 2001, Sitch et al. 2003). Yes, the fire module is important for simulating disturbances, but we do not modify parameters in the fire module or alter the process representation in this paper. The GlobFIRM module requires much needed improvements or replacement with another fire module. The GlobFIRM module clearly underestimates burned area, both regionally and globally. The assessment of GlobFIRM, relative to other fire modules and datasets, are already reported elsewhere (Poulter et al. 2015, doi:10.1002/2013GB004655; Hantson et al. 2020, doi: 10.5194/gmd-13-3299-2020).

Reviewer_1_Specific_Comment_006: L219–221. Presumably the fractional area abandoned/logged goes entirely to the youngest element within the youngest age class (f0,0, following your notation in Eq. 4), is this correct?

> Yes, correct.

Clarify. Also, does it mean that the model assumes that all recently disturbed areas have similar structure of survivors? This may be fine for abandoned and clear-cut logging, but not very appropriate for fires and selective logging.

> The model does not assume the structure of survivor trees. The structure of the abandoned/logged/burned area that goes into the youngest element is determined by the underlying process. For example, if wood harvest is prescribed to an area, but the demand for harvest biomass is satisfied before all biomass is removed, then there will be 'survivor' trees on the youngest element stand. If a fire occurs on a stand, but the fire does not burn all the PFTs, then there will be survivor PFTs on the stand.

Reviewer_1_Specific_Comment_007: Section 3.1. Are there allometric equations that relates carbon stocks, vegetation height and stem number density in LPJ? I wonder if this could explain the consistently lower stem densities, and if the biomass distribution across size would look more/less similar to the plot data.

> Yes, there are space filling 'packing' constraints on stem density, based on allometric rules for size/height of PFTs. Yes, it could help explain the lower densities in LPJ-wsl v2.0 relative to the FIA plot data. Moreover, LPJ-wsl v2.0 does not represent vertical complexity, such as understory growth, which would increase stem density.

Reviewer_1_Specific_Comment_008: Section 3.1.2. I may be missing something here, but I cannot see which ecological processes are affected by choosing equal or unequal age classes. It almost reads like the only difference between the two simulations is how results are reported, please clarify the mechanistic differences between the two approaches. Also, as a point for discussion, it would be nice if the authors provided some insight of which approach is recommended.

> We clarified in Section "2.2.1 An age-based model of ecosystems – sub-grid-cell patch dynamics" as below
>
> ".. 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. 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. 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_1_Specific_Comment_009: L436–440. These results are a bit expected because recently disturbed patches are more dynamic, so having finer bins for young age-classes makes sense to me. But it is also unclear is the effect of different binning strategies on the final results.

The line reference (L 436-440) was in regards to the emergent pattern in the decline in NEP with age of stand. It is generally expected that NEP declines with increasing age, yes. However, we did not expect to find such consistent patterns between NEP and stand age. We clarify as below,

"The binning strategy is likely not a determinant of this pattern between NEP and stand age, which is evident in Figure 3 for both age-class setups. In this regard, we care less about the binning strategy and more that the emergent pattern is reflective of simulated model dynamics. This emergent pattern could lend itself to observational constraints if similar emergent patterns can be derived from forest inventory data in the future."

Reviewer_1_Specific_Comment_010: Section 3.2.2. Is a recovery of NEP in 5–6 years more reasonable than 30 years? I don't see why, this needs some independent evidence from observations. Also, some clarification is needed to explain why Rh is consistently higher in the no-age simulation. Shouldn't the stand-scale mortality (and turnover) be the same in both cases, and the only difference be how mortality (and turnover) are applied?

Agreed, we state throughout that future work requires additional benchmarking or data assimilation to align model processes with observational patterns.

After a disturbance event, Rh is consistently higher in the no-age simulation, yes. We try to explain the mechanisms that results in this model artefact in the aforementioned Section 3.2.2. Note that mortality and turnover are left unchanged in the model; these processes are the same for all model setups (no-age, equal-bin and unequal-bin setup). The processes of mortality and turnover, among all other processes, act on the state variables themselves.

Reviewer_1_Specific_Comment_011: L518–519. I agree with the authors on the need of more targeted simulation experiments, but if some of the variables mentioned are available from the LPJ-wsl output, then the authors could check the results to see if some hypothesized mechanisms could be ruled out.

More simulations could help explain the fire-age zonal patterns, yes. Ideally, we first would want to make sure that the fire module aligns with burned area observations. We think such investigation is beyond the scope of the current work, and leave it simply as an open question for future investigation.

Reviewer_1_Specific_Comment_012: L647. This would account for only part of the uncertainty. Parameter and process uncertainty in most models can be quite large.

Correct. The statistical model would be emulating a model defined by a specific set of parameters and processes. In an ideal world, the statistical parameters for climate sensitivity and stand age would be constrained by uncertainty simulations, bounded to realistic parameter values.

Reviewer_1_Specific_Comment_013: L688–690. It may be worth mentioning that this size distribution may vary across regions (e.g., Espírito-Santo et al., 2014) and even within region depending on abiotic factors (e.g., Asner et al. 2013 which the authors already cite).

We agree with the recommendation and rephrased as below. (bold for emphasis, here only).

"The distribution of forest gaps also has a predictable power-law relationship with size of the gap (Asner et al. 2013), **which can be allowed to vary across and within regions (Asner et al. 2013, Espírito-Santo et al. 2014),** and this fact lends itself well for representing gaps within the framework of the current age-module."

Reviewer_1_Specific_Comment_014: L700. It makes sense to end the text with a paragraph about future developments, but the current one is vague. Which specific features could be implemented and which ones should be priority?

We agreed that we could do better to prioritize model improvements for the readers. The text in Section 4.3 has been updated accordingly. The beginning of the section now starts as below, with added text to support the suggestions.

"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. ..."

Minor comments

Reviewer_1_Minor_Comment_001: L23. Explicitly say which latitudinal band has the lower age.

Edited accordingly.

Reviewer_1_Minor_Comment_002: L24. Land use change and land management were. . .

Edited accordingly.

Reviewer_1_Minor_Comment_003: L25. Does −21 yr correspond to both temperate and tropical areas? Clarify.

Yes, the difference (-21 yr) corresponds to both temperate and tropical zonal bands. Edited accordingly.

Reviewer_1_Minor_Comment_004: L81. “is” instead of “was”?

Edited accordingly.

Reviewer_1_Minor_Comment_005: L98. This sentence could be dropped, considering that version control software has been around for a very long time.

We agree. Edited accordingly.

Reviewer_1_Minor_Comment_006: L125. I don’t see a strong reason to use both patch and age-class throughout the text. It makes sense to keep the explanation here but use a single term thereafter.

We agreed, we now think use of the term ‘patch’ causes unnecessary confusion. We replaced all instances of ‘patch’ with ‘age-class’ throughout.

Reviewer_1_Minor_Comment_007: Eq. 4. Isn’t the fw,n(t + 1) term a form of fusion? I guess this depends on how independent the different elements within age-class are.

Yes, this is fusion or ‘merging’. We added clarifying text to the Section 2.2.1 to explicitly say that we also merge patches, but we do not merge along axes of similarity.

“... We also merge age-classes, but we do not employ merging rules along arbitrary axes of similarity. We fix the number of age-classes *a priori*, similar to LPJ-GUESS in that there is a maximum number of age-classes. Instead of forced merging to reduce computational burden (as in ED2), a fraction of the age-class always transitions to an older state, and a fractional area can transition and merge with the next oldest age-class. ...”

Reviewer_1_Minor_Comment_008: L175–187. Is there any reason why some of the fractional areas are fw,n and others are Fw,n? If not, then use a single notation.

The text was changed to reflect single elements, fw,n

Reviewer_1_Minor_Comment_009: Also, in Eq. (5), is it correct to say that F 0 totalj (t) = Ftotalj (t) − fw,j (t)?

The meaning of the Reviewer’s comment is unclear.

We edited Eq #5 to show that the sum of fractional areas for all age classes and age widths equals F_total

Reviewer_1_Minor_Comment_010: L202. Rewrite this sentence. Conceptually yes, the approach does seem to avoid dilution, but no example from actual model simulations was provided. Also showing that this approach works in LPJ-wsl is different than saying that the age-class/agewidth approach solves the dilution issue. I am not even sure this is an issue with other models or the default LPJ, are there examples of this happening from the literature or in other LPJ simulations that the authors carried out?

> There are no other examples showing this issue in the literature. We conduct a single-pixel idealized simulation to show this effect directly. In the Panel for Veg Carbon in Figure 5, the post-disturbance biomass in the no_age simulation is diluted. This is the extreme scenario for a single stand, which can be thought of as a simulation within only 1 age class.
>
> When averaging two numbers, the mean will always be less than the maximum value, by definition. The average over a vector of carbon densities ($C\ m^{-2}$), which takes into account the contributing fractional areas, will give a mean carbon density that will always be lower than the maximum carbon density in the vector. Hence a dilution of the densities will always occur. The VTFT method tries to reduce this effect. Absent computational constraints, we could represent every land fraction separately and avoid dilution.

Reviewer_1_Minor_Comment_011: L223. “to” instead of “->”

> Edited accordingly.

Reviewer_1_Minor_Comment_012: L233–235. This assumption seems counter-intuitive at least in the tropics, where young secondary forests have high deforestation rates (e.g., Nunes et al. 2020; Wang et al. 2020).

> We agreed. We changed the text to read “This rule will always result in greater land-to-atmosphere fluxes than if rules were employed that allowed younger age-classes to be preferentially deforested.”

Reviewer_1_Minor_Comment_013: L235. At least for me, this seems the opposite of a conservative estimate.

> Agreed, we corrected the text as above.

Reviewer_1_Minor_Comment_014: L262. “were” instead of “was”

> Edited accordingly.

Reviewer_1_Minor_Comment_015: L275. Because readers may not be familiar with FIA plots, include the total plot area and the minimum DBH measured over the entire plot area. Also add the metric equivalents for all diameter references.

We edited and reword the text accordingly. For clarity, it now reads as below:

"... The FIA plot level data are composed of 4 circular sub-plot sample areas (168 m^2), wherein attributes of all trees with Diameter at Breast Height (DBH) ≥ 5.0 inches (12.7 cm) diameter are recorded. ..."

Reviewer_1_Minor_Comment_016: L293. Is the 5% based on any real mechanism?

No, it is a simple way of maintaining fractional areas in every age-class for every year of the simulation. If we did not prescribe disturbance (5% annual clearing), then might not have a distribution of age-classes within a grid-cell. Alternately, we might have a situation where young age-classes are only present once during the simulation, which could occur during dry or wet years.

Reviewer_1_Minor_Comment_017: L306. "Data" instead of "Date"

Edited accordingly.

Reviewer_1_Minor_Comment_018: L375. This seems a software-specific remark, mention and cite the software.

If the line reference above is correct (L 375), then the text refers to statistical modeling, which is not software-specific.

Reviewer_1_Minor_Comment_019: L434. Clarify this text. What is the field-based evidence, and whether the results are consistent with the evidence in a quantitative or qualitative manner (from reading the text it looks like it is the latter).

We edited the sentence as below for clarity. (bold is for emphasis)

"... The age-class module **qualitatively** demonstrates NEP-age relationships..."

Reviewer_1_Minor_Comment_020: L477. What are the differences in GPP?

Within the paper, we focus on differences in NPP as opposed to GPP, which is less certain. NPP is much more easily constrained by observations of changes in biomass.

Reviewer_1_Minor_Comment_021: L484. "(?), perhaps not" is confusing.

We agreed and removed the referenced text.

Reviewer_1_Minor_Comment_022: L489. Isn't it possible to retrieve the soil moisture as a function of age from the LPJ-wsl output? I had understood that soil moisture was solved independently for each age class.

Soil moisture is solved independently for each age-class, yes. Although we output many state variables by age-class, we currently do not have soil moisture as an output by ageclass. We think we understand the Reviewer's point. Such output could be beneficial to a focal analysis or further development of the fire module.

Regarding the context where soil moisture is mentioned in the text, the point we make is that the difference in fire fluxes between the S_{no_age} and S_{age} simulations are probably less to do with soil moisture and more to do with simulating biomass heterogeneity within a grid-cell. After all, each age-class within a grid-cell receives the same exact climate inputs (precipitation, temperature). If it is hot and dry in one age-class, it will typically be hot and dry in all age-classes within a grid-cell.

Reviewer_1_Minor_Comment_023: L493. True, but the apparent large difference for other terms may be just because the scales for most variables do not go to zero in Figure 6. In relative terms they may be comparable to the changes in NEE.

The y-axes are all the same units. Although they are displayed on different scales, the fact that values do not go to zero does not play a role in our interpretation, nor does the relative difference among the state variables. The absolute difference is what matters in this context. It is relevant that there are compensating fluxes from Fire and Rh in the S_{no_age} and S_{age} simulations which contribute to give a similar NEE value.

For clarification -- The compensating fluxes are driven by differences in the distribution of carbon among pools. When we include age-classes in the simulation and see little to no change in global NEE, someone might conclude that there is no important effect of demography. Arguably, carbon stocks in different pools (live vegetation, litter, soil) is easier to benchmark than carbon fluxes from fire or heterotrophic respiration. The differences in the component fluxes and corresponding source stocks are indeed large.

Reviewer_1_Minor_Comment_024: L518. "drier" instead of "dryer".

Edited accordingly.

Reviewer_1_Minor_Comment_025: L549. The central South America looks as strong as the central USA.

Edited accordingly. The precipitation effect generally tracks semi-arid regions, which was a good sanity check.

Reviewer_1_Minor_Comment_026: L610. Including age dynamics is important, but this is not a novel concept, so it would be nice to put this paragraph into perspective with previous efforts.

We revised sentences in the introduction that puts our work into better context, stating that there are existing models that simulate ecosystem demography.

Reviewer_1_Minor_Comment_027: Fig. 2. In case B, shouldn't 0.25 be in the 2nd row of the 3rd column, with a zero at the 1st row? Also, can logging be applied to other age-classes or just the last one? If multiple classes can be disturbed, then it may be worth showing such example too (or replacing the single-patch disturbance with a multi-patch disturbance example).

Reviewer_1_Minor_Comment_028: Fig. 4. It would be interesting to compare these trajectories for the two age-class approaches (equal bins, unequal bins).

Reviewer_1_Minor_Comment_029: Fig. 9. These results are a bit surprising given that boreal forests burn frequently. Could this be caused by the zonal averaging, which puts drylands and savannas together with low-disturbance forests in tropical and temperate zones (but not so much in the boreal zone)?

References

Asner G. P., Kellner J. R., Kennedy-Bowdoin T, et al. 2013. Forest canopy gap distributions in the southern Peruvian Amazon. PLoS One, 8: 1–10. doi:10.1371/journal.pone.0060875.

Espírito-Santo F. D. B., Gloor M., et al. 2014. Size and frequency of natural forest disturbances and the Amazon forest carbon balance. Nat. Commun., 5: 3434. doi:10.1038/ncomms4434.

Fisher, R. A., Koven, C. D., et al.: Vegetation demographics in Earth system models: a review of progress and priorities, Glob. Change Biol., 24, 35–54, doi:10.1111/gcb.13910, 2018. Nunes, S., Oliveira, L., Siqueira, J., Morton, D. C., and Souza, C. M.: Unmasking secondary vegetation dynamics in the Brazilian Amazon, Environ. Res. Lett., 15, 034 057, doi:10.1088/1748-9326/ab76db, 2020.

Wang, Y., Ziv, G., Adami, M., Almeida, C. A., Antunes, J. F. G., Coutinho, A. C., Esquerdo, J. C. D. M., Gomes, A. R., and Galbraith, D.: Upturn in secondary forest clearing buffers primary forest loss in the Brazilian Amazon, Nat. Sustain., 3, 290–295, doi:10.1038/s41893-019-0470-4, 2020.