**REVIEWER REPORT #1**

**General Comment from Reviewer 1 and 2:** (Reviewer #1) Previous points that did not seem to be addressed, at least based on the response letter. (Reviewer #2) * I noticed that responses to Reviewer_1_Minor_Comment_027, 028 and 029, and Reviewer_2_General_Comment_006 and 007 are missing.

An honest oversight. We provide responses to the previous reviewer comments as below.

**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).

In Fig. 2, Case B, the oldest age class (last column of matrix) only has 1 element (1 row). As such, the position of the element in the matrix does not change, so 0.25 does not increment to the next row, which is Null.

Logging can be applied to any age class. We specify in the text the there is a priority rule for logging/harvest. Simulated wood harvest tries to meet the prescribed demand (as harvested biomass according to LUHv2) by harvesting biomass from the oldest to the youngest age class. It is possible that all age classes are logged (harvested) if the demand exceeds biomass on all age classes. A sentence in the Section 2.2.2 sub-section LUCLM trys to clarify this, "(2b) wood harvest (i.e., biomass harvest) also occurs in the ranking of oldest to youngest age class until two conditions are met. Timber harvest occurs on each age class until a prescribed harvest mass or harvest area is met."

We do not entirely understand what the reviewer means by multi-patch disturbance? If the reviewer refers to situations where fire and wood harvest (for example) both happen in the same year among different age classes, then yes, this can happen. We thought this point was extremely relevant to add comments on and we added the text below during discussion of age class integration with fire and land use.

"The disturbance processes of simulated fire, land use change, and land management can occur on multiple age classes at a time. That is, these processes are related but independent. For instance, fire can occur independently on each age class, and each age class would have its own independent estimate of the probability of fire. Wood harvest occurs first on the oldest age class and progressively harvests younger age classes until two demands are met (harvest area and harvest biomass); described in detail in the relevant section below. Clearly, each process influences the other as logging or fire both remove biomass that could be potential fuel for a future fire or biomass for a future harvest. These relationships are not evaluated here, but are noted for its potential importance. Below, we describe in detail the integration of the age-class module with those the two prominent forms of disturbance: fire, land use change, and land management."

**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).

We agreed and added an additional Panel to Figure 4 to show NPP and Rh vs Age (linear across time as opposed to age class codes) for both Equal-bin and Unequal-bin age class simulations. We thank the reviewer for the recommendation and realize that this version of the figures is more informative and presents the comparisons more succinctly.

**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)?

I would not use the term 'frequent' for fires at boreal latitudes, but perhaps that's relative to a certain baseline? It is unclear what the reviewer is referring to as a surprise finding. If the reviewer is commenting on the small trend due to fire at boreal latitudes (Fig. 9), then I would not expect this to be too surprising. The fire season is short at high latitudes, relative to the tropics for instance. Most fire simulation models underestimate burned area at boreal latitudes, including GlobFIRM, but even still, the burned area fraction at tropical and temperate latitudes is greater than at boreal latitudes based on multiple datasets (*see* Figure 1 and Figure 5 *in* Hanson et al. 2020 https://doi.org/10.5194/gmd-13-3299-2020).

In regards to zonal averaging of multiple biomes within zonal bands: Yes, good point. I would also expect that the simulated trend in zonal age due to fire and/or land use change is certainly driven by some biomes over others. Certainly, the fire return interval in tropical rainforests is less common than in tropical savannahs, which often have annual fires. The intent of the analysis, summarized in Figure 9, is to evaluate general time-varying zonal patterns of ecosystem age as a function of fire and land use, with emphasis here on simulated age distributions. The simulated statistical trends among distinct zonal bands are informative, even if the analysis lacks attribution to specific biomes. That said, this point by the reviewer deserves a follow up focal analysis on causative drivers of trends in age distributions in specific biomes. The comment is well noted.

**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?

We compared the FIA data against idealized Regional simulations results, aggregated to the USFS Divisions. We realized that the sub-section headers probably added to the confusion here because the subsequent section describes the regional simulations. We combined the (former) sections 2.3.2 and 2.3.3 and renamed as, "2.3.2 Examining age dynamics: qualitative evaluation of regional simulations against U.S. Forest Inventory Analysis (FIA) data to assess simulated changes in stand structure and ecosystem function".

The regional simulations do not prescribe logging or simulate fire. Instead, we impose a 5% fractional disturbance, as described in the (new) section 2.3.2, sub-heading 'Regional Simulations'. We describe our reasoning for this idealized simulation as, "The intent of the setup was to ensure that each grid cell maintained fractional area in every age class for each year of the simulation and avoided situations in which age classes were only present in 'bad years', or when growing conditions were poor."

For clarity, we rephrased the text describing exclusion of some FIA plot data as follows, "We only included plots ... with no history of major disturbance, stocking, or logging (DSTRBCD=0, TRTCD1=0), *which could alter natural patterns of tree density versus age and confound the comparison to simulated data.*" (italics for emphasis here only).

**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.

We agreed overall and have updated Figures 3 and 4. We thank the reviewer for the recommendation and realize that this version of the figures is more informative and presents the comparisons more succinctly.

**Reviewer_1_Specific_Comment_002:** I understand that the authors did not use these options, but in this case either delete the text in line 101–102, or at least make it clear upfront that these will not be discussed.

We added text as below to clarify, italics for emphasis here only.

"The model known as LPJ-wsl v2.0 is based on LPJmL v3.0, but includes modifications to managed lands that now includes modeling gross land cover transitions, forest age cohorts, and also a modification that include permafrost and wetland methane*; the permafrost and wetland modules were not used in this study."*

**Reviewer_1_Specific_Comment_006:** I think the authors' response provides some useful examples on how survivorship is handled in the model, and could be incorporated in the text.

We added the following text (italics for emphasis here only) into Section 2.2.2 when discussing survivor trees.

"It is possible to have so called 'survivor' trees on the youngest age class that then skews the age-height distribution of the age class. *The model does not assume any structure of survivor trees. Instead, survivor trees occur as a function of the underlying process. For example, 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. Both fire and wood harvest (below) are simulated based on fractional area, and it is the fractional area, specifically, that gets reset to a young age class.*"

**Reviewer_1_Specific_Comment_007:** Same, the authors could complement the manuscript with the text from their response, as it at least provides some additional possible explanations for the observed discrepancies.

We integrated our prior response into the text as below. Bold to emphasize changes to the text, here only.

"However, LPJ-wsl v2.0 is a big-leaf, single-canopy model **that include space-filling 'packing' constraints on stem density, based on allometric rules for size and height of PFTs**. Also the model does not represent multiple PFT cohorts in an age class, or more simply, it does not represent vertical heterogeneity **such as understory growth that would otherwise increase stem density**. As such, and under the current model architecture and associated assumptions, the **exact** cause of the mis-match is unclear."

**Reviewer_1_Specific_Comment_010:** I could have been clearer in my previous iteration. The opening sentence of section 3.2.2 (Line 479–480 of the current version) implies that a 30-year recovery window for NEP is unrealistic, but the authors did not provide any independent observation to support that a 5-year recovery window is more adequate. I suggest to drop the first sentence, or to provide some support to this claim.

In Section 3.2.2 we added text throughout to clarify when we refer to gridcell-level fluxes. Perhaps 'unrealistic' is not the term we should have used. We changed the context of the text in reference to refer to 'model artefact' because the simulated pattern of 30-years of NEP recovery is purely a consequence of model construct. What we mean by that is that a patch-based model is more like reality, where the full 'grid' of space is an explicit representation of unique patches of ecosystem. It could be representation like Reimann's sum, where each bar is a patch and it approximates a fraction of unique ecosystem states, itself an interesting way to conceptualize patch models. Whether or not the recovery times themselves are accurate is less concerning at this point. The growth rates and recovery trajectories will have to be optimized, ideally, to observed patterns, but this is beyond the scope of this paper. We added the following text to the end of section 3.2.2 to clarify,

"The VTFT module also uses the mean-individual approximation but stand dynamics are always allowed to occur in natural progression and the relatively small age widths (10-years) ensure that stand age dynamics (NEP-age trajectories in Figures 3 and 4) most evident in the first 50 years are discretely modeled. To reiterate, we think that the simulated flux dynamics in the no-age simulation is a pure model artefact. What we mean by that is that a patch-based (age class) model is more like reality, where the full 'grid' of space is an explicit representation of unique patches of ecosystem. Whether or not the recovery times themselves are accurate (30-years vs 5-years) is less concerning at this point. The growth rates and recovery trajectories will likely have to be optimized, ideally, to observed patterns, but this is beyond the scope of this paper."

**Reviewer_1_Specific_Comment_012:** The authors could add a sentence or two in the manuscript to address the need to account for parameter and process uncertainties.

We added the following paragraph at the end of the Discussion Section "4.3 Forecasting Demographic Effects with a Simplified Statistical Model":

"A last note on emulators. Useful statistical emulators have fidelity to the underlying process model, but such emulators often cannot address uncertainty from parameter values that are often fixed in the underlying process or uncertainty in process representation. In an ideal world, the statistical parameters for climate sensitivity and stand age, for instance, would be constrained by uncertainty simulations that are themselves bounded to a realistic range of parameter values in the process model (Zaehle et al. 2005) and alternate representations of ecosystem processes (Forkel et al. 2016)."

**REVIEWER REPORT #2**

**L 22:** How many are "a few"? Please, be specific.

We replaced 'few' with 'two'.

**L 91:** Substitute "prior studies" for "previous studies" to avoid repetition.

We edited the text accordingly.

**L 92:** Move "are presented" to the end of the sentence for better understanding.

We edited the text accordingly.

**L 120:** Is mortality also formulated as dependent on tree size? This formulation will link to competition by space among individuals.

> Mortality is not dependent on size, per se, but there is a space-filling constraint that results in mortality via density reduction. We clarified as below.
>
> Mortality occurs as in the original version of LPJ, "...a result of light competition, low growth efficiency, a negative annual carbon balance, heat stress, or when PFT bioclimatic limits are exceeded for a period of time" (Sitch et al. 2003).

**L 134 and over the text:** Replace ageclass and ageclasses with age-class and age-classes or with age class and age classes if preferred.

> We changed all instances to "age class(es)", but kept "age-class(es)" when used as a adjective.

**L 346:** Correct "by products".

> We specifically mean 'byproducts' (also written as 'by-products') to refer to "secondary products made in the manufacture of something else."

**L 425:** Change pft for PFT.

> We edited the text accordingly.

**L 431:** Remove "it" from the sentence as: "Changes in species composition over time do occur and can add…"

> We edited the text accordingly.

**L 645-647:** This sentence is a bit difficult to read. Consider break it or rephrase.

> We agreed and rephrased the text in reference as below.
>
> " In some geographic locations, it is certainly possible that our wood harvest priority rules (defined by harvesting oldest age class first) might lead to simulated stand ages that are younger than observed stand ages if other harvest rules were applied in real life. For example, if there are a mandates to preserve old-growth forests, then logging might preferentially occur on young or mid-aged forests, leaving older age class forests unharvested."

**Table 2:** Great summary of the study!

> Thanks!

**Figure 3:** Please, check the y-axes for Stem Density and NEP. They do not match between unequal age widths and 10-yrs age widths. Labels from the y-axes of the right column plots could also be removed.

For clarity, we redesigned Figure 3 to be a joint figure, which puts

**Figure 10:** Plots have different dimensions. Please, expand the width of the right column plots and if possible, make the axes comparable. The legend is duplicated and could be placed at the bottom of the figure.

We edited the figure accordingly and kept a single legend.