Thank you for your review and your comments that help to clarify the manuscript. Below we duplicate your comments (**bold**) and respond to them point-by-point (*italics*) followed by modifications that will be adopted in the revised manuscript.

1. The text raises two important expectations that were not fulfilled in the results. First is the issue of land-atmosphere interactions. I would argue that it is conventional to think of land-atmosphere interactions in terms of energy and water budgets, and indeed some authors (e.g., Santanello Jr et al. 2018) define land-atmosphere interactions exclusively in terms of energy and water budgets. Yet the current manuscript contains no results related to energy and water budgets. Personally, I think that such results would be an interesting addition to the paper. Without such an addition, I think the authors need to change the text to more specifically refer to carbon fluxes and budgets.

Thank you for pointing this out. We adapted the text in several places to not raise the expectation of presenting results related to energy and water budgets (just for variables influencing energy and water budgets, in particular LAI). We further tested the response of evapotranspiration (ET) and we attached a figure showing the change in NRMSE for ET to the review responses (Fig. R1). Figure R1 shows that the shape of the change in NRMSE is comparable to that for LAI, however, the magnitude is smaller. We thus feel that the ET plot would not add much information and decided not to show the figure in the manuscript.

Examples of sentences changed in the text: The second and third sentence of the abstract now state

Forest age-structures in turn influence biophysical and biogeochemical interactions of the vegetation with the atmosphere key land surface processes, such as photosynthesis and thus the carbon cycle. Yet, many dynamic global vegetation models (DGVMs), including those used as land surface models (LSMs) in Earth system models (ESMs), do not account for subgrid forest age structures, despite being used to investigate land-use effects on the global carbon budget or simulating land-atmosphere interactions-biogeochemical responses to climate change.

The second sentence of the summary and outlook section now states:

JSBACH4–FF allows land-atmosphere interactions key land surface processes to be simulated in dependence of forest age and, simultaneously, to trace the exact forest age, enabling the which is a precondition for any implementation of age-based forest management schemes in JSBACH4–FF.

Second is the issue of forest management. The authors repeatedly state (see P11,L10-11 for one example) that an advantage of the model is in simulating different forest management scenarios. However, this is not exploited in the paper (I know that the forest management scenarios in Fig. 7 are different, but so is the climate, so one cannot isolate the effect of the management scenario). Why not illustrate the power of the new modelling approach by running different forest management scenarios for a single grid cell?

The main purpose of the paper was the description of the implementation of forest age-classes in JSBACH4 and the presentation of the applied new approach for this introduction of forest age-classes. The possibility to implement different forest management scenarios is one important motivation for this model development, but neither the only motivation, nor a focus of our paper and we consider running the model for several more forest management scenarios beyond the scope of this study. In order to not suggest the focus of our study is studying effects of various forest management scenarios we adapted the text in several places. However, we would like to note that technically we in fact do compare two different forest management scenarios, namely the average harvest in the PFT simulation and the harvest of the oldest forest area in the age-class simulations (under the same climate).

#### Examples for adapted text passages - line 8-9 first page:

Our scheme combines The first being a computationally efficient age-dependent simulation of all relevant processes, such as photosynthesis and respiration, without loosing the information about using a restricted number of age-classes. The second being the tracking of the exact forest age, which is a prerequisite for the any implementation of age-based forest management.

#### First sentence of the results and discussions section:

Having forest age-classes in JSBACH4–FF facilitates a finer discretisation in each grid-cell and enables the is a precondition for any implementation of age-based forest management.

## Additionally we inserted a statement on the two applied forest management schemes in the Section on harvest maps (now 2.3.4):

In different simulation types – with or without age-classes – the same harvest maps were used, but different forest management schemes were applied. In simulations with age-classes, a clear-cut according to the fractions in the harvest map was taken from the oldest age-class. In the simulation without age-classes, the PFT simulation, we used the same harvest fractions as in the simulations with age-classes, but harvest was applied as done in JSBACH3 (Reick et al., 2013), i.e. by diluting the wood carbon of the harvested PFT tile.

2. Some of the methods were not adequately justified. First, I am concerned that the initial condition is unrealistic. Why did simulations begin in 1860 from bare ground rather than from a spin-up? Of course, the 1860 initial condition would more realistically be represented by many forested areas.

Thanks for pointing this out, we have now added our reasoning to start from scratch in 1860:

Simulations started in 1860 from scratch, i.e. with empty vegetation carbon stocks, and were run up to 2010. Empty carbon stocks are a simplification used in the absence of global knowledge on the state of the forest in 1860, but have no influence on our results, since in simulations with JSBACH4 (4.20p7) LAI, GPP and AGB only depend on the age since the last clearing event, not on the history before that. The starting date of 1860 was chosen such that it covers at least one full cycle of regrowth, as the oldest age resolved in the simulations matches that of the observation-based data (Poulter et al., 2018).

Note that the used JSBACH version has no influence of soil carbon and nutrient state on vegetation growth, which could be influenced by the history prior to the last clearing event,.

## Second, I could not determine from the paper how one goes from the 2010 age-class distribution to time-dependent (1860-2010) harvest rates. This procedure should be described.

We did outline the derivation of the annual harvest maps in 2.4.3. (now 2.3.4), and described the procedure in detail in the supplementary material S2.For the procedure deriving the harvest rates we politely point the referee to the listing in S2.

Third, it seems like there is an inconsistency between the definition of the model's PFTs (tropical evergreen and deciduous, extratropical evergreen and deciduous) and the Poulter et al. PFTs (broadleaf evergreen and deciduous, needleleaf evergreen and deciduous). What is the correspondence?

We added the mapping formula to 2.4.3 were we previously only stated that the Poulter et al. (2018) PFTs were mapped to JSBACH's PFT cover fractions.

#### Part of the edited text in 2.4.3:

The map by Poulter et al. (2018) provides a grid with 0.5 ° resolution of the global forest age distribution of 4 forest PFTs(needleleaf evergreen and deciduous four forest PFTs: needleleaf evergreen (NE) and needleleaf deciduous (DE), as well as broadleaf evergreen and deciduous) on a grid with 0.5resolution (BE) and broadleaf deciduous (BD). The map uses a discretisation into 15 age-classes, covering 10 years each, with the last class containing all area with an age >140 years. In a pre-processing step, the map was remapped to T63 using the conservative remapping operator of the CDOs. Subsequently, the PFTs from the map were scaled to area sums of the two evergreen and the area sums of the two deciduous PFTs from Poulter et al. (2018) were used to derive the age-class maps for JSBACH's PFT cover fractions. From these scaled evergreen and deciduous PFTs, respectively, following Eq. 6

3. The comparison to observations can be made more substantial. RMSE is a helpful statistic, but I wonder what is being missed by only considering this statistic. For example, I wonder what can be learned from Taylor diagrams? I am certainly not asking that the paper include Taylor diagrams for every variable, but rather such diagrams could be analyzed in a preliminary analysis and the most exciting ones presented in the paper or supporting information.

Following this suggestion we performed an analysis using Taylor diagrams comparing the PFT simulation, *i.e.* the simulation without age-classes with the IAS11 simulation, *i.e.* the age-class set up used in the more detailed comparison in the manuscript. We included the Taylor diagrams in the supplementary and added a note in the methods section. However, since we found that the Taylor diagrams do not allow new, important conclusions to be drawn beyond the NRMSE comparison, we did not include further analysis of the Taylor plots in the results section.

#### Text added to Section 2.2 (previously 2.2)

## In addition, we created Taylor diagrams for each variable, season and region (see Figures S5.5–S5.11 in the supplementary).

4. There are some problems with the interpretation of the results. First, I think it misses the point to repeatedly state that the new model is better. Rather, the fundamental result is that new model tends to reduce GPP and LAI relative to the old model. The new model is better because the old model was biased high. If the old model had been unbiased, then the new model would have been biased low. Alternatively, suppose that there is another modeling group excited by this study, and that that modeling group has a model that is biased low. Then implementation of this scheme would probably make that model worse.

We agree that our implementation of age-classes in a model with a low bias would probably make such a model even worse and adapted the manuscript to stress that introducing the age-classes in a model which is biased low would lead to an increase in the comparison error. We also agree that results closer to observations could just be a consequence of compensating for a high bias (for whatever reason that high bias may have existed) in the old model and now distinguish between model improvement in terms of quantitative results (which could be disputed) and model improvement in terms of inclusion of processes known to exist in reality (in the latter respect the new model is clearly "better"). Nevertheless, we would like to point out that spatially explicit comparisons of the results from the PFT simulation and observation-based data ("OBS-PFT" in Figs. S4.2–S4.4, column 2) indicate several areas of underestimation (red) and of overestimation (blue) for all variables, thus the old model was not merely biased high.

In addition, we have tried to understand if the high bias in the old model is due to not including ageclasses or due to other processes. We found that indeed part of the high bias stems from missing an adequate representation of regrowth: we attached a figure showing the change in model bias per mean age to the review responses (Fig.R2). Figure R2 shows that the change in model bias decreases with forest age, indicating that the error reduction happens where the old model was biased high due to not considering forest age.

In the results and discussions section looking at the benefit of having age-classes (3.2) we inserted the following summary and caveat:

In summary, simulations using age-classes led to a decrease in the simulated GPP, LAI and AGB values due to their non-linear increase with a saturation for older ages. This caused a decrease in the NRMSE<sub>Max-Min</sub> in areas where the PFT simulation was biased high and an increase in the NRMSE<sub>Max-Min</sub> in areas where the PFT simulation was biased low. Thus, if such a forest age-structure would be implemented in a DGVM being predominately biased low, the difference to the observation-based data could increase.

Second, I think that more care needs to be taken in the interpretation of Figure 6. While the curves in panels a-c are decreasing, the authors do not quantitatively support their assertion that the curves are decreasing exponentially (and not, say, quadratically). Exponential fits should be done and the quality of the fits should be analyzed if the authors want to assert that the declines are exponential. Related to this, the assertion that there is "no offset" in panel d is unsupported. A linear fit should be done, and analysis of the residuals would inform whether there is an offset.

Concerning the shape of the curves in panels a-c: We eliminate statements about the shape of the curve, since this is not relevant for our conclusions. The second last sentence of the abstract, for example, now states:

The comparisons show differences exponentially decreasing with the decreasing differences and increasing computation costs with an increasing number of distinguished age-classes and linearly increasing computation costs.

Concerning computing time (panel d): The pre-last paragraph of the Evaluation section (3.1) now reads:

Comparisons of required CPU times show a linear near-linear increase with an increased number of ageclasses (Fig. 6d) and neither a difference between the two age distribution schemes, nor an striking offset as compared to the PFT simulation. This behaviourA near-linear increase with an increased number of age-classes was expected, since the processes requiring most of the computing time, such as the calculation of photosynthesis, carbon allocation and respiration, are conducted on the age-classes. The absence of an striking offset comparing the PFT simulation with the age-class simulations indicates that the introduced organisational overhead on the PFT level in simulations with age-classes is not substantial, i.e. tracing of the exact forest age and redistributions of area fractions and other state variables among tiles, is not dominating the computation times.

5. In Section 3.3, note that much of the discussion is also relevant to cohort-based models (or at least the ED family). The ED approach involves discretization of a partial differential equation (equation 5 in Moorcroft et al. 2001), and thus there are again questions of the optimal number of age bins, whether the bins should be of different or equal sizes, and criteria for merging.

Thank you for this comment.

**Technical corrections** 

P1, L8: do you mean "simulation" rather than "implementation"? This paper, of course, deals with the simulations rather than actual implementations of forest management.

We actually meant implementation. We split and rephrased the sentence, now stating:

In this paper we present a new scheme to introduce forest age-classes in hierarchical tile-based DGVMs combining benefits of recently applied approaches. Our scheme combines The first being a computationally efficient age-dependent simulation of all relevant processes, such as photosynthesis and respiration, without loosing the information about using a restricted number of age-classes. The second being the tracking of the exact forest age, which is a prerequisite for the any implementation of age-based forest management.

## P1, L9: not clear what "hierarchy" is being referred to here

Thank you. We edited this sentence.

This combination is achieved by using the tile-hierarchy to track the area fraction for each age on an aggregated plant functional type level, whilst simulating the relevant processes for a set of age-classes.

## P2, L11: replace "extend" with "extent"

Changed accordingly.

P2, L13-16: there are a couple of sentences where a plural verb "are" is used with a singular subject ("one example")

Changed accordingly.

### P3, L5: this sentence seems to have missing words or typos

Thank you, we corrected "extent" to "expand".

# P3, L6: I am comfortable with the idea that this is a frequently applied approach, but do you have evidence that this is the "most frequently" applied approach?

We edited the sentence and now state that it is the more commonly used (of the two recently developed approaches that we present, based on the number of references that we found and list using one or the other approach):

To extent expand tile-based DGVMs to represent subgrid forest age structures, two approaches have recently been developed. The most more frequently applied approach has been to increase the number of tiles in such a way that a certain number of age-classes or structurally similar stands can be distinguished.

### P3, L30: Perhaps instead of "In this paper we try to", use "The objective of this paper is to"

We edited the sentence, it now states:

In this paper we try tobridge the two approaches for extending tile-based DGVMs to represent subgrid forest age in the sense that we present a way to trace the actual age of the forests in a grid-cell despite following the first approach using a restricted number of additional tiles and thus required merges.

### P7, L4: Note that "data" is plural. Hence, "these data".

Changed to "datasets".

Throughout: My sense is that the word "exemplary" is not being used appropriately in the text. Exemplary denotes a particularly good example, whereas I think the authors are oftentimes just referring to an example of the typical sort.

We replaced the three occurrences of exemplary in the manuscript.



Figure R1: Change in NRMSE<sub>Max-Min</sub> when comparing simulated evapotranspiration (ET) of simulations using an increasing number of age-classes to observation based ET data (GLEAM V2A - Miralles et al., 2011). As in the comparison for GPP, LAI and AGB (Supplementary figure S3.1) averaging has been conducted giving equal weights to each of the four seasons.



Figure R2: Change in the difference of the differences of simulated and observed MAM LAI using the PFT and the IAS11 simulation. Dots mark differences in single grid-points at grid-point mean ages, the red line marks the mean change of all dots for each mean-age (rounded to integer years). Note the difference in the y-axis of the two panels.

#### References

Miralles, D. G., et al. "Magnitude and variability of land evaporation and its components at the global scale." (2011).

Change in MAM LAI bias (abs(OBS-PFT)-abs(OBS-IAS11))