Thank you for your review and the suggestions that help to clarify the manuscript. We have duplicated your comments (**bold**) below, each followed by a point-by-point response (*italics*) including the modifications that will be adopted in the revised manuscript.

(1)The description of model "development" part needs to be strengthened. In general, the authors failed to highlight their technical advances or difficulties in contrast to the original model version (or the complexity the development that's achieved) to allow reviewer to appreciate their work. I agree this hierarchical structure in JSBACH is novel and it seems to facilitate the management of which processes to be executed on which level to save computation time (as the authors claimed but is not actually shown). Is this feature out of the development by the authors in this work? Otherwise it seems that the authors just made use of this existing model feature and did some simple configuration changes (mainly the number of age classes and their distribution over age) and they claimed this as a new "development". For "development", I would understand as substantive new features in the model, or improvement in parameterization, or new method for model calibration etc. It seems the only paragraph that's fully dedicated to the "development", or the description of the author's new work is the first paragraph in Section 2.3. All other material in the "Methods" section is devoted to introducing JSBACH model structure (2.1), or existing hierarchical model feature (2.2) or simulation set-up (2.4). With this it's hard to appreciate what's really achieved in this paper in terms of "model development". The whole paper more sounds like testing a configuration of the model in terms of age class and performing some sanity check in GPP, LAI and AGB. That's how I reach the feeling of an interval technical report.

Thank you for this important comment. We rewrote large parts of Section 2.1 and the former Section 2.3 (now 2.2) and adapted Fig.2 and Fig.3 to better explain the newly developed scheme and to better emphasise new model developments. The tile hierarchy did indeed already exist, however, as a purely infrastructural piece of code. For each of the introduced process this infrastructure had to be extended. In particular, we newly introduced the age vector to track the age, and the processes managing the age-classes. These processes were either newly implemented (ageing and harvest) or had to be advanced (disturbances). An additional major technical advancement was to address the new necessity to introduce shifts of area fractions from one AC to another, as well as resulting shifts of forest carbon and the re-determination of other affected state variables. We now also describe these new infrastructural developments.

(2)There is great confusion in this hierarchical model structure and the advantages that the authors claimed to have. If this overall, sharing model "overhead" can really save computation time, we would expect a non-linear relationship in Fig. 6d ? A decreasing amount of extra time used for each unit increase in number of age class should be expected. From this, I don't see the author's claim that such a feature that different age classes share some common "overhead" process to be computationally efficient as being proved.

Thank you for pointing this out. We assume that different concepts got mixed up, being (1) what we called the "organisational overhead", (2) savings of computation time by only introducing a restricted number of age classes, and (3) the potential to save computation time by simulating processes on different levels of the tile-hierarchy. When rewriting Section 2.1 and the former Section 2.3 (now 2.2), as well as upon adding to Fig.2 and Fig.3, we attempted to resolve this confusion. We particularly removed statements targeting point (3) listed above since these are not directly connected to the developments presented in our paper. For clarification: we used "organisational overhead" to refer to the additional computation time required for bookkeeping of the exact forest age and managing merges of fractions of different age-classes.

In addition to rewriting Sections 2.1-2.3 (now 2.1 and 2.2) we explain what we term the "organisational overhead" on its first occurrence in Section 3.1, stating:

The absence of ~~an~~a 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.

(3)Relate to the above point. The authors mentioned throughout the paper the importance of biophysical feedbacks of forest management but nothing of this aspect is shown in the paper. Instead, there is little description on how such processes are simulated in the age-class model structure. The only text I found that gives such similar description seems to be lines 10-13 in Section 2.2 but this is quite vague. The readers are left wandering in what processes belong to "overhead" and which are age-class specific ? For example, how the processes like albedo, energy balance, soil water processes, carbon allocation are simulated ?

When rewriting Section 2.1 and the former Section 2.3 (now 2.2) we attempted to also resolve this point changing the text to make more explicit (1) which processes already were present in the basis version of JSBACH4 and which have been implemented in the course of this study and (2) on which level of the hierarchy the different processes are executed. Regarding raised expectations concerning biogeophysical feedbacks (which was also commented by reviewer 1, comment 1) 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 responses to reviewer 1 (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.

Which of these processes are "overhead" and how flexible they are in terms of being simulated on different levels ? These are critical for the age-class feature to really reflect forest management impacts but are unfortunately little described.

We made several text changes, particularly in the former Section 2.3 (now 2.2), as well as additions to Fig. 2 and Fig. 3, to make more explicit which processes and state variables are located on which level (please also see the response to point 1). Regarding the term "overhead", please also refer to response 2.

(4)The essence of the new age-class feature is to yield lower estimate of LAI, GPP and AGB than the old version. Comparing the overall agreement between the old and new feature with observation is nice but not the most convincing way from my point of view, because the old version can always be adjusted/parameterized to agree with the observation and if this is the case, the new version would show a prevalent low bias.

We agree that the new model version does show a better performance where the old model version was biased high, particularly in several of the regions having young forests. Nevertheless, we would like to point out that the improved model performance is related to not considering forest age (see Fig.R2 in the responses to reviewer 1). Furthermore, 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. Tuning the old version could potentially result in a lower high bias in regions where the forest is young, but this will result in a low bias in regions where the same PFT is mature. So, tuning is not an alternative to including age classes. In order to raise awareness of the general direction of biases (which was also commented on by reviewer 1, comment 4) we inserted the *following summary and caveat in the results and discussions section looking at the benefit of having age-classes (3.2):*

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 $NRMSE_{Max-Min}$ in areas where the PFT simulation was biased high and an increase in the $NRMSE_{Max-Min}$ 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.

What would be nice is to show whether the model improvement is systematically related to the forest age. For example, is the bias or error reduction more pronounced in regions where young forests dominate? What the processes driving such a decrease in simulated LAI, GPP and AGB and how does this relates to the "ageing" process in the model? The author mentioned several times of this "ageing" process but what is it and how does it impact the simulation of these variables? Are examples of new model behaviour related to age-class development is necessary to understand this? Another way to show the influence of this new development is to show its impact on estimated global fluxes, such as land use change emissions as the authors described in the introduction.

For the relation of the improvement in model performance and forest age please refer to Fig. R2 in the responses to reviewer 1. Regarding the simulation of LAI, GPP and AGB: their relation to the "ageing process" stems from LAI, GPP and AGB being simulated separately for each age-class. Due to the non-linear relationship of GPP, LAI and AGB with forest age (Fig.7 of the manuscript) simulations of a mixed aged forest will result in higher values in a mean age forest simulation (PFT) than in a simulation resolving different age-classes and thus leading to independent simulations of LAI, GPP and AGB on tiles representing different forest ages. Regarding the explanation of the ageing process: we added a more detailed description of the process in 2.2:

Ageing The ~~ageing of forests happens annuallyand affects the oldest year in each AC~~newly implemented process of forest "ageing" happens annually: upon ageing each tracked forest fraction gets one year older. Yet, a shift from one age-class to the next age-class only happens for the area of the oldest age ($maxA_{K-1}-1$) of an age-class AC_{K-1}, i.e. only the forest area which upon getting one year older ~~the fraction of forests having age $maxA_{M-1}$~~exceeds the upper age bound $maxA_{K-1}$ of the AC_{K-1} needs to be shifted into AC_K. Thanks to the tracking of the age in the *fractPerAge* vector, the exact area fraction with age $maxA_{K-1}-1$ ~~will shift from AC_{M-1} into AC_M~~is known.

Some minor and editorial comments:

P 3 line 5 : "to extent" -> extend

Changed accordingly.

P4 line 12: "be able to" could be removed.

The sentence has been adapted (see response to the next comment).

P4 line 11: "a dependency of the maximum leaf area index (LAI) on the available leaf carbon ", what do you mean by "available leaf carbon", does it mean existing leaf biomass or NPP that's allocated to leaf? I would think it is rather natural and reasonable that maximum LAI being limited by leaf biomass? How do this feature relate to the age class development ? Is this feature already satisfying for age class structure, or not ?

We agree with the referee that it is natural and reasonable that the maximum LAI, i.e. the LAI that can maximally be reached at the peak of a season, is limited by leaf biomass. However, this is not the case in the standard JSBACH3 (Mauritsen et al., 2019) version and therefore also not in the standard JSBACH4 version. In JSBACH3, the maximum LAI is a PFT-dependent constant, which is why we implemented this dependency in an independent study (Naudts et al., in prep.). We now explicitly stress that this was not the case in JSBACH3 and that it is a precondition for the introduction of our age-classes. We rewrote this part of Section 2.1 now stating:

As an important amendment to the current version (4.20p7)~~ used as basis in this paper has been amended by a dependency of the~~, we ported a new JSBACH3 development, which we implemented in a recent independent study (Naudts et al., in prep.): While previous JSBACH3 versions assumed a PFT-dependent but constant maximum leaf area index (LAI), that is the LAI value that can maximally be reached at the peak of a season, Naudts et al. (in prep.) introduced a dependency of the maximum LAI on the available leaf ~~carbon, which only recently has been implemented in JSBACH3 (Naudts et al., in prep.)~~biomass. Such a dependency is a prerequisite for simulating forest re-growth and thus for the introduction of age-classes.

P6 line 2: is the “git” feature relevant here, it has been mentioned several times including the in the title.

We prefer to keep this information for reproducibility reasons.

P 6 line 4: the upper-bound of what ?

Thank you for pointing this out. We edited this (and other) sentences. This sentence now states:

In addition, ~~the upper-bound of each~~a to be pre-defined upper age bound per age-class AC~~$_M$ (maxA$_M$)~~$_K$ (maxA$_K$) as well as ~~the~~a total maximum age (maxAge) ~~need to be pre-defined~~were introduced.

P 6 line 20: “initiated” can be removed.

This sentence changed upon rewriting of Section 2.3 (now 2.2).

P 6 line 21-22: “which are directed and scheduled on the PFT level but exerted on the ACs ”. I don’t get the meaning, could it be explained in an easier way ?

This sentence changed upon rewriting of Section 2.3 (now 2.2).

P 7 line 4: Some brief introduction on GPP and LAI data is needed. A critical issue here: as far as I understand Tramontana et al. 2016 GPP data does not consider forest age and it’s questionable to use this as a product to evaluate a model with age effect because the age is the key point here. A recent paper by Besnard et al. ERL (https://doi.org/10.1088/1748-9326/aaeaeb) tried to address this but I don’t know whether they have GPP product. Likewise, is the LAI data pure satellite observation?

Thank you for pointing this out. We now briefly introduce GPP and LAI in Section 2.3.1. Furthermore, we added some caveats regarding the observation-based data in Section 3.2. Additionally, we redid Fig.7 in the text using LAI instead of GPP.

Brief introduction in Section 2.2.1 (former 2.3.1):

We used ~~GPP and LAI data for the year~~2010 ~~as derived in Tramontana et al. (2016). This data already had~~MODIS LAI (Myneni et al., 2002) and GPP data obtained from machine learning methods trained on flux-tower measurements (Tramontana et al., 2016).

Added sentences in Section 3.2:

In this context, caveats regarding the observation-based data themselves need to be raised. A known caveat regarding MODIS LAI data is the problem of reflectance saturation in dense canopies making the reflectance insensitive to changes in LAI (Myneni et al., 2002). This problem, which is particularly relevant to the tropical region, could lead to a general high bias of the model compared to the observation-based data. However, since this problem is more typical for denser old grown forests, this high bias would also occur in the simulations with age-classes. Regarding the GPP data from Tramontana et al. (2016), a recent study by Besnard et al. (2018) criticised that the applied empirical upscaling techniques do not directly consider forest age, making it unclear how well they can capture age-related dynamics. In their study, Besnard et al. (2018) advocate the development of alternative global datasets considering forest age as a predictor.

P 8 line 24 : "to be harvested fraction" -> to-be-harvested-fraction ? A noun form should be here but please check.

The sentence was superfluous and has been deleted.

Figure 2: what's the "UML" ?

Now spelled out (Unified Modeling Language).

Figure 3: AC M , I would use AC i , which distinguishes clearly with AC N ,i.e., the former refers to a common AC, while the latter refer to the old-growth AC.

Thank you very much for this suggestion. We updated the figure accordingly (for better readability we used K instead of I). We also replaced all occurrences of AC M in the text by AC K.

Figure 5: Label for vertical axis not consistent with others. Can you use more expressive label, for example, "Normalized RMSE?".

Fig.5 shows the $NRMSE_{Max\text{-}Min}$ for each variable, region and season. Fig.6 and Fig. S3.1 show averages over the seasons or the seasons and the regions, respectively. Therefore, the y-axis of these figures are labeled differently. We added "normalised root mean squared error" to the figure captions.

Equation (2): I would write simply N-1 for the denominator...

In the former EQ.2 (now EQ.4) the denominator should contain the sum of i's (i.e. for N=5: 1+2+3+4 = 10). However, there had been a mistake in the equation (the sum over the i's started with i), which we now corrected.

P12 line 1: "as also discussed" -> as is also discussed

Changed accordingly.

Figure 7, caption: "Stars mark the JJA GPP per age-class", please indicate this is for simulated data.

Changed accordingly.

Could you somehow simply the caption ? It's rather long and almost deters reading.

We tried to shorten the caption, but it still remains long as we prefer to include all the information necessary to read the figure. The caption now reads:

Example grid-points comparing 2001-2010 mean spring leaf area index (MAM LAI) from simulation without (PFT) and with age-classes (IAS11) to observation-based data. The map in the center shows the difference of differences between the observation-based data and the simulations (abs(OBS-PFT)-abs(OBS-IAS11)), i.e. it shows where the results from the simulation with age-classes (IAS11) deviate less (blue) or more (red) from the observation-based data than the PFT simulation results (see also Figs. S4.2-S4.4, column 4). Dashed lines in the map mark the three selected regions (see Table 2). The plots (a-g) show the LAI of selected PFTs (ETD: extratropical deciduous; ETE: extratropical evergreen; TD: tropical deciduous; TE: tropical evergreen) as well as their according area fractions per age-class and per year at the labelled grid-points. Center latitude, longitude and grid-cell cover fraction (cf) of the depicted PFT are indicated. The x-axis reflects the age from 0-151 (purple) with the age-classes (black) indicated at the age centres. The two right y-axes represent the bars: depict are the 2010 area fractions relative to the area of the depicted PFT. Blue bars are per age-class (black y-axes) and depict the fraction of each age-class (i.e. one bar per age-class); the yellow framed purple bars depict the fraction of each age (i.e. one bar per year). The left y-axis depicts the LAI. Stars mark the simulated LAI per age-class, and the lines the LAI of the depicted PFT – blue dashed line: IAS11 simulation, black line: PFT simulation, green line: 2010 value from the observation-based data. Note: 1. The age-class LAI is only depicted for age-classes having non-zero fractional cover over the whole timespan 2001-2010 (this is not the case for the age-classes 9 and 10 in panel c, f and g). 2. Age and age-class fractions of classes 2-8 in panel g are very small and therefore not visible above the x-axis. 3. Since we did not apply any harvest in the final simulation year 2010, the first year and accordingly the youngest age-class are always empty.