Nabel and colleagues presented in this paper a new development made in the land surface model JSBACH that accounts for sub-grid even-aged forest age classes or cohorts that are outcomes of stand-replacing events such as clear-cut or stand-replacing forest fires. Then they performed a simulation from 1860 from scratch, driven by clearcut-like forest harvest data to match a forest age distribution map recently constructed. They compared simulated results with observations for GPP, LAI and aboveground biomass and model results with and without age classes. As shown in their Fig S2 to S4, model simulation with age class exclusively reduced the magnitudes of all three variables considered, which is expected as the original model structure without age class maintains all these variables close to their equilibrium level (quasi old-growth) if the disturbed area is not big enough to yield a clear decrease. Consequently, for regions where original model shows a high bias, the new development brings the simulation closer to the observation.

I believe the model development direction represented in this paper is worth publication. But the evidences and results presented in it are more like an internal technical report rather than a scientific publication (see my first major comment and other minor comments below). Several aspects of the paper have to be improved if it is to be accepted, which I list as below:

(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.

(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.

(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 ? 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.

(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. 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.

Some minor and editorial comments:

- P 3 line 5 : "to extent" -> extend
- P4 line 12: "be able to" could be removed.
- 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 ?
- P6 line 2: is the "git" feature relevant here, it has been mentioned several times including the in the title.
- P 6 line 4: the upper-bound of what ?
- P 6 line 20: "initiated" can be removed.
- 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 ?
- 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 ?

- P 8 line 24 : "to be harvested fraction" -> to-be-harvested-fraction ? A noun form should be here but please check.
- Figure 2: what's the "UML" ?
- 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.
- Figure 5: Label for vertical axis not consistent with others. Can you use more expressive label, for example, "Normalized RMSE?".
- Equation (2): I would write simply N-1 for the denominator…
- P12 line 1: "as also discussed" -> as is also discussed
- Figure 7, caption: "Stars mark the JJA GPP per age-class", please indicate this is for simulated data. Could you somehow simply the caption ? It's rather long and almost deters reading.