

## ***Interactive comment on “Using the International Tree-Ring Data Bank (ITRDB) records as century-long benchmarks for land-surface models” by Jina Jeong et al.***

**Anonymous Referee #1**

Received and published: 18 May 2020

In this manuscript, the authors aim to develop a method for benchmarking land surface models (LSMs) against tree ring data. The paper details the issues with using such data, specifically from the International Tree Ring Database (ITRDB) and describes a method for overcoming the biases in these data. This method is then tested using the ORCHIDEE model with four different levels of complexity. This is a valuable study as tree ring data is a resource largely untapped by the modelling community and which can give us insight into tree growth over long periods of time. I also think that the approach the authors have taken is very robust and should be applied in more model evaluation studies - the careful assessment of the biases in the data as well as the model features needed for a comparison with the data.

C1

I believe that a lot of time and effort has gone into this study, however, I think it needs some editing to make it more clear and convincing.

### **Major comments**

Having read the whole paper, it remains unclear to me if the method presented here actually works and in fact, how can we tell if it does work. I suspect that the answers to these questions lie in figure 4, but this figure is very hard to understand. First of all, it comes before the figures illustrating how each of the benchmarks works, so it is not clear to the reader what the different quantities in the figure are. It would help if figures 5-8 came first. Secondly, I am unsure about the comparison between the model and the BACI data here. As far as I understand, the model is run at different sites than the ones in the BACI data. Why is this and are the sites similar in terms of their climate and stand characteristics? Also, why does this figure only show coniferous sites? Is this an issue with the available BACI data? Would the results look similar for broadleaf deciduous sites?

The description of ORCHIDEE contains a lot of general details that I'm not sure are needed in a journal such as GMD e.g. capacity to be coupled/decoupled, variable grid size etc. On the other hand, I find the details of the model setup somewhat sparse, and these details are needed to understand the model results. A simple solution would be to just take the detailed description of the setup from the supplementary material and add it to the main text.

I'm not sure I understand why the four different models are needed, if the specifically stated purpose of this paper is not to evaluate the model. It adds an extra level of complexity that makes an already long and complicated paper even longer. If the application of the four model versions is insightful, it would help if this was discussed somewhere. From Fig. 9 it looks as if for some benchmarks the differences between sites are bigger than the differences between model versions - is this caused by climate, stand age, stand density?

C2

The paper opens with a relatively long discussion about the issues and biases in using tree ring data to benchmark models. The problem is then that the actual discussion section largely repeats the same arguments. It would be more interesting to see here a discussion about the generality of the benchmarks - can they be used at different sites? With different species? For other models?

#### **Minor comments**

L 83 Is the assumption that forests are unmanaged likely to be correct?

L 84 How was the start year set to match observations? Is this based on inferred tree age from the tree ring data or is there more information on forest age?

L 80 Was there data on N deposition also used as forcing?

(Note line numbering appears to break after 100)

Figure 4 - I'm not sure what 'trend' refers to. The model value for 'young' does not appear to have error bars. Why does the data set have 27 sites and the model is run at 10 sites only? Are all these values for coniferous sites only?

L 130 These benchmarks are discussed earlier but only explained here

L 185 Are the results of the leave-one-out approach shown somewhere?

L209 (I think, p 17) I don't see why the dynamic leaf N in itself would cause a problem, as it is a realistic process. It is much more likely that having a more complex representation of N processes exposes an issue with other parts of the N cycle in the model (e.g soil)

---

Interactive comment on Geosci. Model Dev. Discuss., <https://doi.org/10.5194/gmd-2020-29>, 2020.