

Dear Editor,

Thank you very much for considering our manuscript for publication in *Geoscientific Model Development*. In this revision, we tried to address all your suggestions. Please, see our responses below.

*E1: The authors' responses to the reviewers indicate a nuanced understanding of the tradeoffs related to model complexity, parameterization, etc. In my opinion, their revisions to the main text address most of the reservations that the reviewers raised. However, there are two additional revisions I'd like to see made before publication. First, Reviewer 1 noted that "intermediate" output variables aren't really evaluated, with various aspects of basal area being the only aspects analyzed (Table 3). I agree that it would have been nice to see some evaluation of more basic processes, but I would guess this to be challenging with the forest inventory data. The authors mention that "intermediate" processes have been evaluated previously, so a paragraph or two summarizing these previous results (rather than just directing readers to the papers themselves) would address this concern.*

**R1.** In our previous revision, we already responded to Reviewer 1 that the evaluation of some 'intermediate' output variables (i.e. LAI, GPP, NPP) will be the focus of our future work. We cannot include it here because it has not yet been performed, but the discussion acknowledges this current limitation (L652-654). What we have done in this new revision, is to add a sentence summarizing the evaluation results of soil moisture dynamics, transpiration and plant water status that we conducted using data on experimental plots (L378-390), which is important to sustain the meta-modelling exercise.

*E2. Second, I have reservations about the use of the same data for parameterization and evaluation. The authors raise a fair point: that because they only intend this parameterization to be used for Catalonia, it's valid to use all the available data for parameterization. However, the model is also to be used for a time period (the future) it's not parameterized against. As such, I would like the authors to redo the parameterization using only one of either SNFI2-3 or 3-4. A comparison against the other would then give a sense of how well the model can be applied in novel time periods.*

**R2.** This is a good point. Yes, we used data from the SNFI2-3 and SNFI3-4 together for the calibration of growth, mortality and recruitment parameters. However, while growth parameters were calibrated for Catalonia specifically, mortality and recruitment parameters were derived from data corresponding to the entire Spanish territory. This fact was not completely clear in the previous ms versions, so we included new sentences to clarify the text (L337-339; L413-414). We also included as supplementary tables the growth rates, mortality rates and recruitment probabilities estimated for different SNFI periods and either Catalonia or the whole Spain (Tables B4-B6). We now include an additional figure (Fig. A5) with evaluation results comparing the performance for the SNFI2-3 and SNFI3-4 periods and parameters calibrated using SNFI2-3 and SNFI3-4 altogether, against parameters calibrated using the complementary period. Evaluation results were almost the same for mortality and ingrowth, but we observed larger biases in the case of growth. This happened because observed growth was larger during the SNFI2-3 period than the SNFI3-4. Thus, simulations with growth parameters calibrated using the SNFI2-3 period overestimate growth for SNFI3-4 and, conversely, simulations with growth parameters calibrated using the SNFI3-4 period underestimate growth for SNFI2-3. This difference in growth biases was already seen in Fig. 3 for SNFI2-3 and SNFI3-4 simulations with growth parameters using data from both periods, although it is more apparent in Fig. A5. RMSE values are, however, not much affected by the calibration period, so we consider the performance as similar.

*E3. (1) The change of trait "coordination" to "covariation" is a good one, but it didn't occur everywhere (e.g., line 82).*

**R3.** Corrected everywhere. Thanks.

*E4. Line 138 and caption of Fig. 1: R isn't previously mentioned is thus a bit confusing at first. Please consider explaining that the model is built in R (?) or just removing its mention entirely.*

**R4.** Thanks for noting this. We modified the text to introduce that the user interface is in R language (L137). We also tried to clarify the caption of Fig. 1.

*E5. Table 3 is really busy, making it hard to get a real sense of the different models' performance. Please consider making a simplified graphical version, with the complete table in the Appendix.*

**R5.** Thanks for this advise. Table 3 is now in Appendix E (including both MEDFATE and IPM) and the performance of the MEDFATE is shown in Fig. 3. We also followed the same advice for the new evaluation figure A5.

We also tried to increase the clarity of the text. All modifications are recorded in the version with tracked changes.