Author Responses to Emilie Joetzjer (Referee) (Responses in Blue)

In this analysis, the authors explore FATES's sensibility to parameter uncertainties using mainly observation-based trait and benchmarking the outputs against BCI's forest inventory and eddy covariance data. With large ensemble simulation, using a single PFT (no competition), they first evaluate the model. FATES performs reasonably to represent physiological processes and forest structure. The authors also show the systematic biases of the model and explain the further needed development. Adding competition leads the model to simulate higher productivity and biomass, pointing out that even though multiple PFTs should be a more realistic configuration to represent tropical forest, the "mono-specific" approach performed better. They also show that, with more disturbances, the model favor early successional PFTs (over late successional PFTs). They conclude on the importance to differentiate parameters that are or are not influenced by competition to better quantify the source of uncertainties in VDMs.

I found the paper proposed by Koven et al, fully relevant to the community. Besides, despite the complexity of the model and the exhaustive sensitivity analysis performed, the paper is relatively easy to follow, the model description is crystal clear and the key messages are well highlighted. I particularly appreciated the effort of the authors to explain the sensitivity analysis through the ecological processes undergoing in the model. Therefore I think the paper is ready for publication, and I only have minors comments (or questions) detailed below.

We thank Dr Joetzjer for her careful review and positive assessment of the manuscript.

Comments

In the introduction, it might be nice for non-land surface modelling reader to sum up in a sentence the connections between ESM, LSM and VDM.

We will clarify these distinctions in the revised text.

L163: Is RH calculated by the LSM or by FATES?

As the domain of FATES is only the vegetative part of the ecosystem, we allow RH to be external to FATES and thus calculated by the LSM.

L 282: Units?

M_c is the rate of mortality of crown area of plants in the canopy strata, and is tracked in m^2/ha/year. F_d is unitless, and thus R_d is also in m^2/ha/year. We will clarify this in the revised version.

L315: Can you describe a bit more BCI and BCI data? Also, what are the time period of the eddy fluxes data? Did you recycled the meteorological data from 1986-2017 to force CLM-FATES several hundreds years, or did you took an averaged climate?

We will describe the site more in the revised version. We recycled the meteorological data over the course of the model simulations.

L369 I didn't get why each simulation is initialized with the observed size distribution. I would guess that the size distribution of a population will strongly depends on the PFT definition. Because FATES is able to compute the size distribution, why not start the simulation from scratch?

The initialization of the size distribution is only to accelerate the convergence of the model. We have found that when we initialize the model from bare ground, we arrive at similar size distributions. Because of the large number of ensembles and experiments here, we chose to initialize the size distributions from census rather than bare ground so as to reduce computational costs.

L434 Might be a naive question, but do you cover a large enough range of the observed variability by taking the same number of ensemble for 1PFT, 3PFT and 10PFT?

This is a good question, and why we chose a relatively large number of ensemble members in each of the simulations: by starting with a relatively deep sampling of the trait space in the 1PFT simulations, we don't expect a qualitatively different behavior due only to that when using larger a number of PFTs. We do sample more parts of the trait space when using a larger number of PFTs and the same number of ensemble members; however, we note that the effect of this larger sampling is actually to decrease the width of the distributions across ensemble members, due to the competitive dynamics described in the manuscript. Thus the more thorough sampling of the trait space itself is unlikely to contribute strongly to that signal.

L530 While I think it's interesting to show how insensitive is FATE to the "host model" by comparing CLM-FATE and ELM-FATE, I wonder how relevant this analysis is. Indeed, both model are forced by the same atmospheric conditions, and as pointed by the author, the only difference reside in the soil representation. My guess, is that both host models provide enough soil moisture to FATE (because they are likely to share the same soil parameterization and BCI is relatively wet), therefore it is logical that FATES behave similarly. I wonder what is the point of this test.

We agree with both reviewers that the justification for this analysis was not sufficiently made in the submitted draft. We do feel that this is important to include in the manuscript, because this approach of modularizing a distinct component of the land surface into a separate codebase that can be called by multiple LSMs (and ESMs) represents an important strategy in managing process complexity and shifting away from a model-centered approach and towards a hypothesis-centered approach to experimental design in simulating the land surface. It is thus important to document the dynamics of FATES within each of the LSMs it currently works within in order to test the hypothesis that such an approach is valid. We will add further discussion of this point in the revisions, as well as linking to a separate

manuscript by authors RAF and CDK, which explores this point in further detail and is currently in review.

L615 I'm wondering how climate variability can play a role in simulating species coexistence (thus my question on the forcing file).

This is a great question, which we don't get into here. This question is explored more fully using the ED2-hydro model and the same meteorological dataset in Powell et al., (2018, https://doi.org/10.1111/nph.15271).

L758 While I agree with your conclusion, you might want to be a bit more explicit. VDM are certainly one of the way to go in ESM to better quantify how environmental changes will affect ecosystems and the associated feedbacks. In the analysis, you show how large (and unconstrained) can be the effect of competition, and how difficult simulating co-existence is. However, the comparison with the observed data suggest that 1 PFT perform usually better than the simulation that integrate competition (Fig. 9). I think it would be nice to have a sentence bridging complexity (and reliability) vs. simple (and therefore easy-to-tune) models.

In the revised manuscript, we will explore in greater detail these caveats and difficulties associated with using this type of model structure, and the costs and benefits of choosing different positions along this axis of model complexity.

Fig. 6 You might want to give more detail on the units of "Tree number density". Is it a number of tree per ha, per diameter class. Idem on Fig. 13.

These units are in n/ha/cm, with the cm^{-1} referring to the width of the size bins used for plotting the curve. Will add this to the figure description in the revised manuscript.

Fig. 9e Is the log scale necessary?

The choice we faced was either to plot the log of the slope or the angle of the slope, as we didn't want to make the sensitivity asymmetric between high-slope and low-slope parts of the domain. We chose a log scale here, we will explore in revisions whether there is a better approach to showing this relationship.