Wednesday, March 6, 2023

Dear Editor,

The manuscript entitled “Does dynamically modelled leaf area improve predictions of land surface water and carbon fluxes? – Insights into dynamic vegetation modules “ by Westermann et al., focuses on the effects that enabling dynamic Leaf Area estimation can have in terms of land surface model performance. The results based on two land-surface models (ECLand, and Noah-MP) suggest that Leaf Area, carbon and turbulent fluxes, and soil moisture representation does not improve when a dynamic vegetation module is employed. This is an interesting finding to the extent that this result feels counter-intuitive, and can be quite instructive on how employing a “fancier” new model does not necessarily increase model performance. Therein lies the novelty of this work. The authors have put some considerable effort and extensively discuss their results providing both their interperation and literature findings, but the manuscript is overall difficult to follow.

Despite I understand that this work could be of interest to members of the land surface modelling community, I am not convinced that the manuscript in its current form is suitable for publication, and would expect that the authors address the following comments and revise their work, before it can be considered for publication.

Comments:

- I understand that the authors use some existing Plant Functional Type in the respective models and do not tune PFT parameters to fit the dominant or average plant traits of each fluxnet site. However, the authors still perform comparisons and evaluate the models against observational point-level ecosystem-specific data. Is this the case? I find this slightly inconsistent, since plant functional types conceptually are mostly meant to be used in larger-scale simulations representing the average characteristics of a vegetation category, rather than being compared with site-level data.
- If we have a model whose parameters are not set to fit the observations (see above), then why would we necessarily expect that switching on a dynamic vegetation module (which is also unparameterized) should increase model performance? One could argue that solely switching the environmental dependency of LAI on should justify this expectation, but isn’t the environmental dependency of LAI on average also embedded in a prescribed climatology by definition?
- I deeply appreciate that the authors are extensively discussing past research, their interpretations, and their results in full detail, which increases transparency, but the manuscript is overall difficult to read. IT would help a lot if the authors split the results from the discussion points. The manuscript would also need further proof-reading, since one can easily still find mistakes scattered across the text.
- The results suggest that model performance “regarding latent heat flux or soil moisture is independent of how LAI is represented” (Line 380). This is very counter-intuitive, and one would wonder whether this is the case because LAI is equally badly represented in all cases.
- Given that LAI dynamics are the main focus of this study, it is important that the authors describe in more detail the LAI modules of the two models. They start doing so in Line 420 and discuss allocation, senescence etc., but they need to do this in a comprehensive manner in the method section, and not scattered across the text.
- I think it is really important to show in the appendix panels with mean annual LAI, GPP, and NEE climatologies for every site, showing prescribed and model-predicted LAI. This would help a lot the interested reader understand the dynamics at play.
- The selction criteria for the fluxnet sites used seem arbitrary. Why do the authors drop sites of roughly similar aridity index? If anything, including more sites would increase the robustness of their results.
- In my understanding, in the switched-on dynamic vegetation runs LAI is freely estimated by the model and not constrained by some prescribed climatology. Is this indeed the case? From the results (e.g., Fig. 2) it feels as if there are different types of dynamic runs – one for each of the different possible LAI forcings. How is this the case? In my understanding, somehow these prescribed climatologies are still used to

"initiate" the dynamic runs. What does that exactly mean? If this is indeed the case, still LAI is free to evolve, so why do we end up with different model performance in every run, just because initial LAI conditions have been different? If this is not the case and LAI climatologies are somehow fed also into the dynamic runs, then how does it make sense to validate these results against the forcing LAI climatology itself? The authors need to try and clarify their setup more.

- Line 155-156: Was it not technically possible to maintain GPP and NEE estimation in the static Noah-MP runs, despite prescribing LAI? Or is there some other reason?
- Line 165: It is slightly unclear in which occasion ERA5 data are used
- Appendix figures: What do the sim00, sim02 etc. mean?
- Line 255: I would suggest that the reviewers show these results in supplementary material to ensure transparency (or don't mention at all)