

Geosci. Model Dev. Discuss., referee comment RC1  
<https://doi.org/10.5194/gmd-2021-150-RC1>, 2021  
© Author(s) 2021. This work is distributed under  
the Creative Commons Attribution 4.0 License.

## Comment on gmd-2021-150

Anonymous Referee #1

---

Referee comment on "Validation of terrestrial biogeochemistry in CMIP6 Earth system models: a review" by Lynsay Spafford and Andrew H. MacDougall, Geosci. Model Dev. Discuss., <https://doi.org/10.5194/gmd-2021-150-RC1>, 2021

---

This article presents a review of the current practices for validating the terrestrial biogeochemistry in CMIP6 Earth system models. The authors use the literature to show that the terrestrial biogeochemistry is a major source of uncertainty in future climate projections, and that this uncertainty can be linked to model structure. They study how 11 modeling groups participating in the CMIP6 exercise validated the terrestrial biogeochemical cycles in their land surface models and their fully coupled Earth System Models. They analyze the different validations presented by the modeling groups in terms of number of variables validated, quantity represented (for instance GPP), spatial and temporal scales, reference dataset and statistical metrics used. They also present two community methods designed to validate land surface models (ILAMBv2.1) and Earth system models (ESMValTool2.0). They present a critique of the validation approaches and suggest ways forward: mainly developing a standard protocol for validation which could be based on the existing communal software packages ILAMBv2.1 and ESMValTool2.0

The article is very well written, very clear and very detailed. The authors read in detail the articles describing the validation of the CMIP6 models. They give a very detailed and informative description of the techniques used by each modelling group in Annex A. I have a few detailed comments (see below) and only two general comments:

- I agree that the use of communal validation software packages would be highly beneficial to the community. However, these should be used to help understand the behavior of the models and help improve them. But I would be reluctant to see them used as tools to select models allowed to participate in certain exercises. If this was the case it would contradict Lovenduski and Bonan's recommendation to improve process understanding and observation accuracy instead of reducing model spread (line 378).

The authors don't suggest to use these validation softwares as selection tools but it might be implied by their remark about GCP on line 217

- The authors don't mention at all the links between the terrestrial biogeochemical cycles and the hydrological cycle. They mention that the modelers should evaluate the response of the biogeochemical cycles to temperature or the effect of nutrient limitations but never mention the effect of moisture conditions. I think the evaluation of the coupling between water and biogeochemical cycles at the land surface is as much of a concern.

Detailed comments:

L94 I find the figure a bit difficult to understand. An example would help like GPP being the variable corresponding to the left-most bar (if I understand correctly).

Table 6: the authors choose to separate vegetation carbon and vegetation biomass although these two type of data are very much related. I guess this is due to a choice made by iLAMB and ESMValTool. But Saatchi et al for instance, although being a tropical biomass dataset also gives data in terms of aboveground and belowground vegetation carbon mass. I'd suggest adding a comment specifying that these 2 quantities are not independent.

L215: The problem with the quality criteria of the Global Carbon Project is that they neglect the uncertainties related to the atmospheric forcing. As shown by Lawrence et al, 2019 (and also shown in Figure 3 in this paper) these may have a pretty strong impact. GCP switched from the CRU-NCEP forcing to the GSWP3 forcing in 2019, and model results were very different. Modeling teams had to retune their models to fit the criteria. That doesn't mean criteria shouldn't be imposed, it just shows their limit.

L 249: I would argue that the soil moisture control on all the variables of the biogeochemical cycles is as crucial in a climate change perspective as temperature, if not more because more uncertain (for instance the future of peatland/wetland in the high latitudes)

L320-324: I totally agree that site-level evaluation is very important to really understand how a model behaves. This is partly because at site-level, there is usually much more information available: the type of vegetation, the type of soil, the presence of an aquifer, land-use practices (irrigation, multi-cropping) etc. The meteorological forcing is also much more precise. In a global simulation, this type of information comes from global dataset (for instance soil texture, depth to bedrock, vegetation type) or are calculated. And this adds a huge part of uncertainty not related to the model structure but to model dataset. I

think it might be interesting for the readers to get that insight

L419: I probably miss something here but I don't see the link between N and P monitoring and turnover time assessment

Technical comment

L23: I am not a native speaker but I find this sentence strange: "growing season extension in cold-limited regions, enhanced heterotrophic respiration, *and potentially others*, as well as environmental"

L378: I believe it should be "precedence" instead of "president"