Comment on gmd-2021-150

Anonymous Referee #2

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-RC2, 2021

This paper does an excellent job of reviewing the land biogeochemistry subcomponent evaluation approaches conducted by different modeling groups for 16 CMIP6-generation land surface models, both in a coupled and uncoupled context. This is useful since finding all the individual land model evaluation papers can be challenging. While Table 1, reprinted from Arora et al. (2020), provides such references, it lists only 11 models, not the 16 models identified in Figure 1. This paper would better serve the community with a table similar to Table 1 that includes all the models reviewed in the paper.

While the scope of this paper is to review only the land biogeochemistry subcomponents of land surface models, describing differences in hydrology and radiation/energy subcomponents and summarizing the assessments of them would be beneficial. Even if the authors do not wish to expand the scope to include other interacting model subcomponents, I recommend adding more key model configuration characteristics to an updated Table 1 to provide additional context to the summary. For example, the table could identify which models employ a river transport scheme, which models have depth-resolved soil carbon, which have an explicit permafrost representation, which used a prognostic dynamic vegetation scheme, etc.

The authors discuss the use of community-developed model evaluation and benchmarking packages, ESMValTool and ILAMB, which are increasingly being adopted as a means of standardizing model evaluation metrics and observationally constrained reference datasets. The wide variety of variables, datasets, and metrics employed in assessing model performance by different modeling groups makes direct comparison across models challenging.

Section 4.1, Variable Choice, begins with, "A comprehensive validation of a process-based model should include all simulated interacting variables for which a reliable empirical reference is available." However, the discussion that follows includes only biogeochemistry variables and there is little acknowledgement of the important interactions with water and

energy variables. It may be useful to add a sentence or two that explicitly mentions the interdependence of and need to co-evaluate carbon, water, and energy cycles.

The authors do not describe the difference between metrics useful for evaluating offline models versus those that can be used for fully coupled models. Fully coupled models should exhibit the same statistical variability over decadal time scales as indicated by observational data, but the observed timing of ENSO and other drivers of climate variability are not reproduced in fully coupled models. Thus, metrics that assess biases or RMSE of time series model output should not be used when evaluating fully coupled model output. This discussion of applicable approaches should likely be included in Section 4.3.

The citation to Lawrence and Bonan et al. (2019) in line 371 should likely be Bonan et al. (2019).

On line 378, "president" should be "presedence".

On line 426, "in junction" likely should be "in conjunction".

Developing a standard validation protocol for model intercomparison activities within CMIP would be useful, and it has been done to some extent for CMIP6 historical land and ocean model performance in comparison with corresponding CMIP5 models in the IPCC AR6 Working Group I report in Figure 5.22, currently accessible at https://www.ipcc.ch/report/ar6/wg1/downloads/report/IPCC_AR6_WGI_Chapter_05.pdf#page=214