Reviewer comments
Author responses

The paper devoted to analysis of quality of terrestrial carbon cycle simulation using Earth System Models from CMIP6. Improvements of CMIP6 models comparing to CMIP5 and empirical datasets are shown. Data compared using a set of statistical parameters and colorful maps. Methods and the aim of the paper are clear.

We thank the reviewer for their comments on our study.

Specific comments

Soil carbon storage, net primary productivity and carbon turnover time were selected as variables responsible for terrestrial soil carbon estimations. According to suggestions NPP related with soil carbon through plant ang root litter (line 30-35), but empirical datasets have negligible correlation between these values (line 458). Pleas, give more attention for the support of your idea on relations of soil carbon and NPP.

We do not *a priori* assume a relationship between soil carbon and NPP, but we do see such a relationship clearly in the CMIP5 and CMIP6 models. Instead, we follow previous studies (Todd-Brown et al., 2013; Koven et al., 2015) in defining an effective turnover time τ_s that ensures that the soil carbon $C_s = R_h \tau_s$ at all times. For the multiannual means considered in this paper, R_h is approximately equal to NPP (because the difference between NPP and R_h, which represents the Net Ecosystem Productivity, is a small fraction of the NPP). We can therefore safely assume that Cs ~ NPP τ_s, which allows us to separate above ground drivers of soil carbon (NPP) from below-ground processes (τ_s). Our analysis makes no other prior assumptions about the extent to which soil carbon is determined by NPP in the models or the observations. We make this clearer in our revised paper by editing the following text (lines 128 to 135):

***"The definition of the effective turnover time** $\tau_s = C_s / R_h$ **ensures that the soil carbon at any one time is given by:** $C_s = R_h \tau_s$. In an unperturbed steady-state (i.e., neglecting disturbances from land-use change, fires, insect outbreaks etc.), there is no net exchange of carbon between land and atmosphere, and **therefore** R_h is equal to **litterfall, known as fallen organic material from plants. When vegetation and soil carbon are close to a steady state, litterfall and** R_h **are also approximately equal to** Net Primary Productivity (NPP), **where NPP is defined as the net carbon assimilated by plants via photosynthesis minus loss due to plant respiration.** In the contemporary period **considered in this study,** R_h **has been found to be well approximated by NPP (Varney et al., 2020). This is because the difference between NPP and** R_h**, which represents the Net Ecosystem Productivity (NEP), is a small fraction of the NPP over the historical period (NPP ~ 60 PgC** yr^{-1}**; NEP ~ 3 PgC** yr^{-1}**).** Therefore, the present day soil carbon can be approximated by:*

$C_s \sim NPP\ \tau_s,$

***to a good accuracy. This allows for** a clean separation of soil carbon variation into the above (NPP) and below (τ_s) ground drivers of soil carbon spatial patterns**, following the approach of previous published studies (Todd-Brown et al., 2013; Koven et al., 2015).**"*

We have also edited the sentence in the Discussion which addresses this issue (line 568):

*"Despite NPP driving the spatial pattern of soil carbon stocks **due to carbon input from vegetation**, a positive correlation ~~is~~ **was** not expected in the real world due to regions with high soil carbon not correlating with regions of high NPP. For example, in the observational derived data soil carbon stocks are greatest in the northern latitudes due to long turnover times in these regions, whereas NPP is lower due to cold temperatures in these regions limiting vegetation growth."*

Carbon turnover time determined as a ratio of carbon amount and heterotrophic respiration. According to presented results soil carbon estimations were improved in CMIP6 comparing CMIP5, but soul carbon turnover time estimations is not good enough. Likely the issue is related with heterotrophic respiration. Could you check the hypothesis and present an analysis of quality of HR simulations?

For the reasons outlined in the previous response, heterotrophic respiration and NPP are very similar on the multiannual timescales considered in this paper. This was shown in our previous paper (Varney et al., 2020), which is now cited in the new text shown above. It is also noted that global total values for heterotrophic respiration (Rh) are presented for both CMIP5 and CMIP6 ESMs, including comparisons with observation, in the Appendix Tables A1 and A2.

Changes in soil carbon storage occurs through changes in fluxes. The accuracy of simulation of carbon fluxes will result in total estimations of soil carbon. You have shown only one flux (NPP) not directly related with soil system and give a complex parameter related with heterotrophic respiration. Is it possible to demonstrate the quality of simulations of carbon fluxes relates with soil system (i.e. heterotrophic respiration, ecosystem respiration, dissolved carbon runoff, decay rate, litterfall, etc)

The global CMIP5 and CMIP6 Earth Systems Models do not yet routinely include dissolved organic carbon (DOC). In any case, reliable global datasets of DOC are not available for model evaluation. Additionally, DOC is known to be relatively small (0.28 ± 0.07 PgC yr^{-1}) compared to the magnitude of NPP (approximately 60 PgC yr^{-1}) on a global scale considered in this study (Nakhavali et al., 2020). As explained above, NPP is a key driver of soil carbon as it provides the litterfall input, and we do have access to global datasets of NPP. Fortunately, NPP, litterfall and heterotrophic respiration are all very similar on the multiannual timescales considered in this paper (because vegetation and soil carbon are close to a steady state on those timescales). We make this clearer by adding the following sentence between lines 128 and 135 (also included above):

*"In an unperturbed steady-state (i.e., neglecting disturbances from land-use change, fires, insect outbreaks etc.), there is no net exchange of carbon between land and atmosphere, and **therefore** R_h is equal to **litterfall, known as fallen organic material from plants. When vegetation and soil carbon are close to a steady state, litterfall and** R_h **are also approximately equal to** Net Primary Productivity (NPP), **where NPP is defined as the net carbon assimilated by plants via photosynthesis minus loss due to plant respiration."***

The paper contains a lot of statistical information about comparison of results from CMIP6/5 ESMs. Total estimations and spatial variability of parameters are shown. But the meaning of obtained estimations and relations with land ecosystem is missed. In the present form the paper is more suitable for Geoscientific Model Development journal where ESM and their characteristics are discussed. Understanding of reasons of ESM errors requires identification of an ecosystem types where highest discrepancies observed. Clear, that highest soil carbon is typical for peatlands. Proper simulation of peatland water, thermal and nutrient regime will giver more impact to the global carbon estimations than for other ecosystems. I suggest to emphasize the role of ecosystems in soil carbon formation and discuss the errors and improvements of ESMs not only at global scale but at ecosystem scale too.

We maintain that this study is very appropriate for publication in Biogeosciences, as it relates to previous studies in this journal (e.g., Todd-Brown et al. 2013, *Causes of variation in soil carbon simulations from CMIP5 Earth system models and comparison with observations*), and is clearly relevant to biogeochemical cycling. We note that Reviewer 1 also suggests that our paper is a good fit to the journal.

Additionally, we follow the reviewer's suggestion to discuss ecosystem types associated with the representation of soil processes, where we have added the following text to section 4.2.2 (line 601) on the important role of peatlands:

"***Different processes control soil carbon formation in different ecosystems, including stabilisation by clay particles, transformation by microbes, nitrogen and phosphorous availability, etc. (Witzgall et al. 2021). In the present study, the largest discrepancies in both soil carbon and turnover times are seen in permafrost and peatland areas (see Fig. 2 and Fig. 7). For example, the west Siberian peatland complex stands out on the majority of the panels in these figures as an area of high model error. This is partly because the soil carbon turnover times and quantities are largest in these regions, but also partly due to the specific controlling processes in these ecosystems. A key part of soil carbon development in permafrost regions is the fact that organic material can be preserved in frozen soil, including via cryoturbation and yedoma deposits, which have not yet been thoroughly represented in models (Beer, 2016; Zhu et al., 2016). There are a variety of other factors, such as plants storing significantly more of their carbon below ground instead of above ground in cold climates, and recalcitrant vegetation such as mosses, which are not represented in most ESMs (Sulman et al., 2021). Peatland formation is controlled primarily by waterlogging, which reduces oxygen available for decomposition, but there are a huge number of additional physical and biogeochemical feedbacks that take place (Waddington et al. 2015). These kinds of small-scale processes and inhomogeneities are difficult to resolve in global models with ~100km2 grid cells, and this***

should be weighed up against their relative impact on global carbon budgets when considering including these processes in ESMs. However, it is suggested that the large-scale discrepancies such as in the permafrost and large peatland areas can and should be resolved in future model versions."