Response to reviewers for BG-2022-92: "Controls of intermodel uncertainty in land carbon sink projections"

Referee #2:

This manuscript takes 11 ESMs from the CMIP6 archive and attempts to unravel the root causes of the uncertainty in the NBP simulated by each model over the (roughly) 2 deg warming scenario. The models' sensitivity to CO2 (through the 1 percent runs) and to temperature and soil moisture are investigated for both short and long timescales. I found the paper figures to be generally well designed (though see my comment about Fig 3) but the text could be confusing at times. There is a lot of rather convoluted steps/arguments in producing the T/SM/sT/sSM metrics and it sometimes was hard to understand exactly what they were telling me about the models. I think the paper is publishable, but needs revisions for clarity. An obvious target for clarity/context would be to discuss the results of this work in the context of previous efforts as discussed in the introduction (principally Arora et al. 2020). I found myself comparing this work to that paper and not understanding why they differed strongly in some cases.

We appreciate the overall positive evaluation of our manuscript. We rephrased some sentences in the text to clarify what do sT and sSM represent and to explain their effect as drivers of differences in cumulative NBP.

Our analysis is complementary to that of Arora et al. (2020) by focusing on the relevance of soil moisture and temperature for land carbon sink projections. Moreover, we provide more regional insights and address projections under a low emission scenario. Thus, most of our results are not directly comparable to those of Arora et al. (2020). Nevertheless, when possible, we don't find strong differences with the insights from Arora et al. (2020) on why each model projects either a relatively low or high land C sink. For example, the contribution of sCO_2 ($\Delta GPP/c'$) is similar in both studies for explaining differences across models. We believe there was a slight misunderstanding of two statements in our manuscript, as discussed in our reply to other comments below.

Main comments:

Fig 3: I found this to be a strange figure. So if a model has a positive correlation it counts towards the blue end of the colour scheme, whereby if it is negative it counts towards the red. However this seems to have no consideration of how positive or negative a model was. I think it would treat a model that is 0.9 the same as one that is 0.0009, which seems to be a bit too ambiguous. Also what if all 11 models are +0.0001 vs. all models are >0.9, as is they would appear the same in this figure but arguably the situation where all 11 are >0.9 is more interesting than the situation where the models are rather ambiguous (close to 0). I would suggest reconsidering this figure.

We modified the figure and now show the ensemble mean correlation divided by the standard deviation to better illustrate the confidence in the sign of the relation and the intermodel discrepancy. The full picture of the actual correlations for each model are provided in Figs. S5 and S6.

Deserts - how did you mask the deserts? I think the grid cell sizes of the ESM precludes removing many deserts, e.g. the Atacama. Instead it seems like only a few were removed

(Sahara, around Middle East, and Gobi) but I am not sure why those made the cut but not, for example the deserts of western Australia or the US SW. What impact does it have keeping them in? Greenland makes sense since there is no vegetation at all but some of the world's dry regions have been fingered as influential in the global C cycle (e.g. Ahlström et al. 2015), so exactly where masking applies could have impact I would assume.

Thanks for pointing this out. Masked grid cells correspond to those with observed annual mean precipitation below 100 mm based on the GSWP3 dataset for the period 1985–2014. The resolution of the grid indeed plays a role in which grid cells are masked. The transition between semi-arid and desert regions likely depends on the resolution of each model. Precisely because of this, we regridded all models to the same grid and omitted the same grid cells to make them more comparable.

We now include in the supplement a Figure like Fig. 8 when not masking desert grid cells.

Fire - Fire is mentioned on line 335 but ignored otherwise, why? I see you mention which models do fire in Table S1.

We also refer to model differences in disturbance fluxes in the manuscript and in Fig. S2, but now clarify that this is strongly related to fire. We now also clarify that sT and sSM may indirectly account for model differences in fire emissions.

Smaller comments:

Line 44: 'with drought-related observed decreasing trends in leaf area' consider rewording, confusing.

We rephrased the sentence for clarity.

Line 103: CanESM has an implict N cycle (empirical downregulation scheme see Arora and Scinocca 2016)

Thanks for pointing this out. Our original statement is based on Table 2 from Arora et al. (2020), where it is mentioned that CanESM5 has no representation of the N cycle.

L 160: This explanation is confusing. Perhaps spell it out in a bit more detail.

We expanded the text for clarity.

L 182: Why 22.5 degrees and not 25 or 30 or some other number? It just seems awfully precise for a seemingly arbitrary limit.

Tropical regions are defined as those with latitudes lower than 23.43 degrees. Given that are employed grid has a resolution of 2.5 degrees, we choose the threshold of 22.5 degrees as the nearest one to the definition of tropical regions. We now also include in the supplement an additional figure when using a threshold of 30 degrees to distinguish which months are considered as the warm season.

L 235: remind reader that both use CLM?

We now also mention here that both models have CLM5 as their land surface model.

L 290: 'underestimation of the land carbon sink modelled by NorESM2-LM and CanESM5,' where is this shown? I can't seem to see any figure where CanESM5 sticks out with an underestimation of the land C sink but it is mentioned here and line 357, indeed in Figure 1 it seems to have one of the highest cumulative NBP. What am I missing?

This is referring to the comparison between the regression estimate and the projected land C sink from the ESM as shown in Fig. 6. In this paragraph it is not about comparing the land C sink across models. We expanded the text to clarify this point.

CanESM5: Other papers (Arora et al. 2020) have suggested that CanESM5 has the largest land C uptake (at least for the 4X CO2 simulations) so it is surprising that it is suggested to be underestimated for the land C sink. Can you clarify how the same model appears to be on the low/high end depending on the analysis? I realize these are different scenarios but I would have assumed high CO2 sensitivity would follow in both (but be exaggerated in the 4XCO2 run), but I don't see high CO2 sens in Fig 7. I assume I missed something here as you mention the Arora et al. paper in the intro but don't return to place your results in context of those other works.

There is a slight misunderstanding. We do not suggest that CanESM5 underestimates the land C sink compared to other models. As mentioned in the reply above, in that part of the manuscript we just note that the regression estimate of the land C sink for CanESM5 is lower (underestimated) than the actual land C sink projected by CanESM5.

It is indeed the case that CanESM5 shows a very high land C sink compared to other models, particularly while CO_2 concentrations and temperatures are rising until approximately the year 2070 as shown in Fig. 1. Northern mid and high latitudes strongly contribute to the high land C sink in CanESM5 as shown in Fig. 2. Our findings suggest that the predominant reason for this high land C sink in CanESM5 is its sensitivity of NBP to T (sT in Fig. 7), rather than its sCO_2. It is clear that CanESM5 has a more positive correlation of NBP and T over the Northern hemisphere compared to other models (Fig. S5), whereas its sCO_2 is not at the high end of the ensemble (Fig. S4). These findings are consistent with the results from Fig. 8 in Arora et al. (2020) where CUE_{Δ} (somewhat related to sT) contributes more than $\Delta GPP/c'$ (related to sCO_2) to the high land C sink projected by CanESM5.

L 352: And assumedly many of them use Nemo for their ocean so model commonalities are not just atm/land. Never mind all who use Farquar photosynthesis etc

This is true. However, the purpose of our statement is just to note that the two models with the lowest projected land C sink share the same atmospheric model, and the two with the highest projected sink share the same land model. We rephrased the sentence to clarify this.

L 378: 'Outperfom' seems out of place, consider swapping it out with something like 'be more important than'

Thanks for the suggestion. We rephrased the sentence.

Lit cited:

Ahlström, A., Raupach, M. R., Schurgers, G., Smith, B., Arneth, A., Jung, M., Reichstein, M., Canadell, J. G., Friedlingstein, P., Jain, A. K., Kato, E., Poulter, B., Sitch, S., Stocker, B. D., Viovy, N., Wang, Y. P., Wiltshire, A., Zaehle, S., and Zeng, N.: The dominant role of semi-arid ecosystems in the trend and variability of the land CO2 sink, Science, 348, 895–899, 2015.
Arora, V. K., Katavouta, A., Williams, R. G., Jones, C. D., Brovkin, V., Friedlingstein, P., Schwinger, J., Bopp, L., Boucher, O., Cadule, P., Chamberlain, M. A., Christian, J. R., Delire, C., Fisher, R. A., Hajima, T., Ilyina, T., Joetzjer, E., Kawamiya, M., Koven, C. D., Krasting, J. P., Law, R. M., Lawrence, D. M., Lenton, A., Lindsay, K., Pongratz, J., Raddatz, T., Séférian, R., Tachiiri, K., Tjiputra, J. F., Wiltshire, A., Wu, T., and Ziehn, T.: Carbon–concentration and carbon–climate feedbacks in CMIP6 models and their comparison to CMIP5 models, Biogeosciences, 17, 4173–4222, 2020.

Arora, V. K. and Scinocca, J. F.: On constraining the strength of the terrestrial CO2 fertilization effect in an Earth system model, https://doi.org/10.5194/gmd-2015-252, 2016.