

Biogeosciences Discuss., referee comment RC1
<https://doi.org/10.5194/bg-2022-13-RC1>, 2022
© Author(s) 2022. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on bg-2022-13

Andy Pitman (Referee)

Referee comment on "Evaluation of soil carbon simulation in CMIP6 Earth system models"
by Rebecca M. Varney et al., Biogeosciences Discuss.,
<https://doi.org/10.5194/bg-2022-13-RC1>, 2022

This is a useful and reasonably thorough assessment of CMIP6 Earth System Models' ability to simulate soil carbon in comparison to both CMIP5 and observations. The paper is well written, and the subject matter is entirely appropriate for Biogeosciences.

Major comments

Multiple conclusions are reached around CMIP6 being differently constrained in performance to CMIP5. The problem is, both CMIP5 and CMIP6 are ensembles of opportunity and you get to analyse what is there, not what you would like to be there. So, the 11 CMIP6 models are different to the 10 CMIP5 models in many ways. Some CMIP6 models are tweaks on CMIP5, some are significantly updated. There are different climate models, there are models with shared land surface models (CESM2, NorESM2-LM) and I would bet that there are shared modules within some of these land models. Obviously, you cannot compare like-with-like and I do not think there are enough members of the ensemble to suggest testing for independence. Thus, in my view all you can do is be very careful in your discussion and conclusions. For example, line 268 – is this reduction real or a function of the ensemble construction? Similar, lines 514, 580, 551 (and others) feel to me to be important in this context. You cannot "fix" this. But you can highlight it explicitly and thoughtfully in the Discussion and make sure your conclusions are limitations-aware.

A more specific question is should you be showing the ensemble mean or the ensemble median? The median is more insensitive to outliers. Why show the mean?

Abstract – there are a series of really nice results in this paper that are not to be found in the abstract. For example, line 565 might have been shown before for CMIP5 but it is important that it is still there in CMIP6. Lines 601-602 are also important. The abstract basically says CMIP6 is an improvement – but hidden later are some really important conclusions that hint that this improvement might be related to the ensemble or at least

"improvements, but for the wrong physical reasons". I think your paper will have much more impact if some of the important findings are reflected in the abstract.

Line 154 you use *mrsos*. I understand why. This is soil moisture over 10 cm. You use this with 1m soil carbon. At the very least you need to discuss whether this is important to your analysis in a "caveat" section in the Discussion. In effect, you are using a high frequency soil moisture (I accept primarily for patterns) which probably reflects high frequency rainfall more than anything else. I am not proposing you change it, but the implications of this needs to be explained and discussed. Your observed soil moisture is also over a shallow depth I suspect (line 220). Does this matter – perhaps not because mostly you only consider patterns but I was less sure when you look at the constraints later.

Line 591 and whole paragraph. There is an issue hidden in here that might be worth discussing which I think was picked up in papers by Jeff Exbryat. Relations with temperature are anchored in reality – and are relatively easy. A land model should simulate a temperature that reflects observations. In contrast, soil moisture is far more complex and most land models achieve a value of soil moisture that reflects the value that is needed to constrain evaporation. There is, of course, much more to it than this, but soil moisture can vary a lot between land models, while all those land models simulate similar drainage and evaporation. This means that when looking at correlations between *ts* and *theta*, the *theta* is not necessarily translatable across models. So, I think there is much more to it that you hint at in this paragraph and it might be worth teasing that out a little. Specifically, the statement "is likely to be due to key soil processes not being represented" might be true but is not very insightful ... and it might be more to do with the relationships between existing ley soil processes not being blended well.

Finally, I liked your conclusions, but I would like to test a couple of them. Are you *sure* about Conclusions 1 and 2. I mean, are you sure you can conclude this despite the uncertainties associated with the ensemble design, the observed data and so on.

I am therefore going to recommend major revisions but with the note that they are not "major" in the sense that a lot of work needs to be done but I do suggest that resolving 1-5 would substantially improve the paper. I would suggest:

- Add a caveats section to the discussion and focus on the major limitations of your study.
- Check your conclusions and see if there are any minor edits you might want to make given the caveats
- Check over the paper and remove anything that is superfluous (see below)
- Check each section to ensure that the narrative flows.
- Make sure the reader knows the supplementary figures exist more obviously.

Minor comments and pedantry

Line 12 – is this statement right - that there is 2-3 times the amount of carbon in the soil cf. the atmosphere. I would have guessed its closer to an order of magnitude. Roughly 600 Pg C in the atmosphere, 550 in vegetation, 1500 in soils and 1700 in permafrost. Please check.

Your paragraph structure needs some attention. Line 16 starts with real world carbon storage. Half way through you switch to Earth System Models. Similar issues in section 2.3.1. You do this a lot. It makes the narrative disjointed throughout and harder to follow than it should be. A significant re-write is really required focussed on the section structure.

Line 22 – How do you know that the most up to date ESMs make up CMIP6? It is an ensemble of opportunity and not everyone who can participated in CMIP6.

Line 22 – “ensemble known as CMIP6” is mis-stated. The CMIP6 ensemble is far broader than ESMs. Just be careful with the precision of the language.

Line 25 – further to pedantic comments, it is not true that the carbon cycle is fundamental to obtaining accurate future projections. It is, of course, true on long timescales, but it is not true of projections to 2050. Again, be precise in the language and the timescales you are talking about.

Line 45 This paragraph is very confusing. It is trying to signpost what comes but fails to do that. You can just delete it as anyone reading this who cannot find your results needs help.

Line 60 – I am not certain the detail in this paragraph really helps the reader. I mean, I am not sure ... consider the value and whether its essential for the reader.

Line 78 – similar comment to line 60.

Line 110 – why divide by 1E12 ?

Line 129 – I do not think you define the term after NPP

Line 344 – the hatching is invisible on the version of the manuscript I read

Line 510 – the word “may” is incorrect. It does reduce our confidence.

Line 526 – this is a brave statement – adding complexity does not necessarily improve simulations.

Line 621 – present day carbon is not “vital” on all timescales. Check language.

Line 625-6 should be in the acknowledgment.

Figure 1-3,6,7 – I am not sure you can change this but the ensemble average of two ensembles of opportunity do not really tell us that much. Maybe you have no choice, but I’d have liked to have seen the individual models. Now, later I find you do this in the Supplementary Figures (ok – that makes sense) but unless I missed it, you do not refer to these anywhere. At the very least, you should flag the existence of these figures in the legends to Figures 1-3,6,7 and in each section when you introduce the ensemble figure you should point to the existence of the individual models.

Figure 1-3,6,7 – again. Take a look at Figure A3. You present the ensemble average of this figure in the main text. Is that “legitimate(as distinct from commonly done). You are averaging together large changes of opposing signs.

Figure 4 – Ahhh, Taylor diagrams. Do you think your audience will understand them? In my experience, the only people who really understand them are those who have personally created and analysed them. For this journal I would suggest you consider their suitability.

Figure 9 – there are two shading intensities I think but its almost invisible on my printout.