Response to referee #2

We thank the reviewer for taking the time to review our manuscript and provide constructive comments. Below we address the reviewer's comments point by point. We add our replies in italics and highlight suggested modifications in the manuscript in red.

We further would like to acknowledge that while revising the manuscript, we noticed two minor errors in the original submission, which we have now fixed:

1. Figure 1d): For two models (ISBA-CTRIP and VISIT) the monthly CSoil pool was shown. We corrected the figure and now show the annual average for both models.

2. Figure 7c) and Table B3: We accidentally chose the wrong time period for the annual fire emissions and updated the figures with the corrected time period (2003-2018 for the TRENDY models and the observation data). We accordingly corrected Table B3. While the annual values for fire emissions and the associated uncertainty almost stay the same, the correlation coefficients between the TRENDY models and the observed data changed so that now all models are positively correlated with either of the two observation datasets on both timescales. We changed the text accordingly to:

   On monthly timescales, four TRENDY models capture some features in the variability in the satellite derived observations with either weak (ISBA-CTRIP and VISIT: both datasets; JSBACH: GFED4s), moderate (JSBACH: CAMS-GFAS; CLASS-CTEM: CAMS-GFAS) or high (CLASS-CTEM: GFED4s) significant correlation coefficients. The remaining models do not show a significant relationship to either of the datasets. Aggregated to annual values, the TRENDY models generally underestimated the fire CO_2 emissions and did not always capture the variability in, or timing of, extreme fire years (see fig. 7c). CLASS-CTEM, JSBACH and ISBA-CTRIP captured some features of the variability in the satellite derived observations. CLASS-CTEM is moderately correlated with both datasets ($r>0.5$), ISBA-CTRIP shows a significant moderate correlation with the GFED4s dataset only and JSBACH is highly correlated with both datasets ($r>0.7$; see table B3). The remaining models are not significantly linked the satellite derived observations.

General comments

This manuscript analyzes how the Australian carbon cycle is simulated by 13 vegetation models that exhibit large differences in their outputs and behavior. It's an interesting and comprehensive analysis, looking at factors such as carbon residence time, land cover differences, and fire.

We thank the reviewer for the positive assessment of our work.

There are some minor weaknesses. The text is unclear in some spots, and some of the figures could be improved or re-thought.

We thank the reviewer for their input. As described in the response to reviewer 1, we tried to incorporate suggestions made but none of them added insight beyond the original figures presented so we have kept the original presentation. It is of course challenging with 13 models and multiple carbon axes of interest, but by separating by experiment (CO2, climate), examining as a function of time, landcover, etc, we have elicited important insights into model behaviour/skill across the Australian continent. Importantly, this comprehensive assessment was currently lacking. Finally, we have improved the highlighted sentences as suggested (see below).

Finally, I didn't see anything about data or code availability; this is critical for transparency and reproducibility.

We're unclear what we have missed here and are happy to adjust the text as the Reviewer deems necessary. Our data availability section listed a link to our github repository with all the analysis code and we put all the links to the datasets used to evaluate the models.

In summary, this is a well-done and interesting analysis that needs moderate revisions in many areas for clarity and concision.

Specific comments

1. Lines 18-20: somewhat confusing

   We thank the reviewer for the comment and rephrased the text to:
   In addition, we find that differences in the timing of simulated phenology and fire dynamics are associated with differences in simulated/prescribed vegetation cover and process representation. We further find model disagreement in simulated vegetation carbon, phenology and apparent carbon residence time, indicating that the models have different types and coverage of vegetation across Australia (whether prescribed or emergent).

2. 54 and 465: perhaps start new paragraph

   We updated the manuscript accordingly.

3. 80: what is NATT? Not defined yet

   We thank the reviewer and updated the manuscript:
   We remapped all model outputs and satellite datasets (see below) to a common 0.5°C grid using first order conservative regridding (except for the comparison with the data over the North Australian Tropical Transect (NATT)).

4. 177-179: unnecessary? Perhaps move to figure caption

   We thank the reviewer for the suggestion. However, as this is a description of results, we feel it does belong in the results section rather than in the figure caption.

5. Figure 1: why does panel a show the ensemble mean and spread but other panels show individual models? Perhaps add green mean to panels b-d?

   Since all models vary around zero for NBP, it is hard to distinguish individual models over a time period of 118 years when each model is shown. We therefore included the spread to show the uncertainty associated with simulated NBP. In contrast, the remaining three variables have lower interannual variability and showing individual models allows to see the different trajectories they take over time. In summary, while we're not advocates for the use of the ensemble mean, we do so in panel a as we think it adds value to the interpretation for the reader.

6. Not sure how useful figure 7a is

   We aimed to explore the possible link between ENSO and fire regimes in Australia with panel a (as described in l. 247-253). We do agree that if we were to delete a panel, we would delete panel (a) but on balance we have decided to leave it in. We would be open to further advice from the editor on this of course.

7. Figures 8 and 9 are interesting!

   We thank the reviewer for their positive assessment.