The authors investigate the future ocean carbon sink in CMIP6 models under several SSP scenarios. They quantify the uncertainty in the future sink as a function of model, internal and scenario uncertainty for the global and regional scales. They calculate time of emergence for the forced signal to emerge from internal variability. They find that the future ocean carbon sink is most uncertain in regions of currently highest flux.

The methods used are solid and, with a few minor exceptions, adequately explained. Reasonable assumptions are made. The Supplementary provides useful additional information.

The paper is a contribution to the literature on the CMIP6 models. Most of the calculations done here have been done before for CMIP5 models in several papers, so this is a useful update. It is appropriate for ESD readership. Conclusions are justified by the analysis

Many thanks for the careful and thorough review of the work and your positive reception. We appreciate the constructive comments and suggestions.

Major Comments

On the whole, the paper lacks depth and clarity in the discussion of mechanisms on the ocean carbon sink and how it should evolve in the future. The references to the literature are also somewhat sparse. I encourage the authors to review some more of the literature and to add more mechanistic discussion and connection to previous studies. To do so will make the paper a more useful contribution. Possibilities include Crisp et al 2022 in Reviews of Geophysics, Ridge and McKinley 2021 in Biogeosciences, Hauck et al. 2020 in Frontiers, McKinley et al. 2020 in AGU Advances, Bushinsky et al 2019 in GBC, Schwinger and Tjiputra 2018 in GRL.

Thank you for the recommendations. We have revisited/reviewed the suggested literature and changes have been made to the manuscript in order to embed the study in the previous literature and provide a more mechanistic discussion where possible. We agree that a more mechanistic approach would make this paper a better contribution to the community and changes have been made to the manuscript to serve that purpose. The following is a summary of the updates to both clarify the novelty of the study as well as include a mechanistic overview:

1. Wider citation of previous relevant studies throughout the paper to show how our results support or are different from previous studies and to emphasize places of novelty in our study. The added citations include; Lorenz, (1969); Somerville, (1987); Sarmiento et al., (1998); Lovenduski et al., (2007); Graven et al., (2012); Fay & McKinley, (2013); Bopp et al., (2015); Landschützer et al., (2015); Frölicher et al., (2015) & (2016); Wang et al., (2016); McKinley et al., (2017); Riahi et al., (2017); Toyama et al., (2017); Bushinsky et al., (2019); Schlunegger et al., (2019) & (2020); McKinley et al. (2020); Hauk et al., (2020); Ridge and McKinley, (2021); Terhaar et al., (2021); Crisp et al. (2022);

2. A new paragraph on detectability using time of emergence reviewing previous studies in the introduction.

3. Overview of mechanisms associated with surface-depth connectivity and how they affect the regional patterns of uptake in the historical period and the future.

4. Discussion of the mechanisms responsible for patterns seen in the maps of the sources of uncertainty; especially large model uncertainty in the Southern Ocean and association between regions of large scenario uncertainty and large uptake.

5. Discussion of the mechanisms driving patterns seen in the time of emergence (TOE), and the association of scenario TOE differences with scenario uncertainty (sensitivity to forcing).

6. Two sections were added to the Supplement. Section S5 provides the details of our analysis for diagnosing the “highly active regions” using a new metric. Section S4 details the test done to compare the multi-model trend with that of the observations.

I also recommend changing away from the ”hotspots” terminology. For the carbon sink, this term is often used for very small regions, such as western boundary current mode water formation regions. To use this for all of the Southern Ocean, North Atlantic, etc. is also just not a very good choice of words also since these are large, basin scale regions not “spots”.

(Also answering the following comments moved from minor comments below:)

pg 8 line 7-8: the “hotspots” terminology is too vague, making it hard for the reader to follow.

Pg 20, line 53: “mostly in a few hotspot regions” suggests a few small spots when in fact the ocean carbon sink is diffuse and occurring really everywhere (see figure 1c). See Major Comment.

We can see your point here. We have used “highly active regions” instead in the revised version. Moreover, a new section is added to the Supplement (section S3) to clarify how we classify the “highly active regions” - 70 percent of the total sink occurring in less than 40 percent of the global ocean.

Minor

Pg 2, line 46. Bushinsky et al. 2019 should be added to this list

Done!

Pg 3, line 84. ESMs are based in fundamental equation such a Navier Stokes. Yes, there are many details that differ, but there is also a lot of basis in physics! This statement suggests that models are much more of a potpourri than they actually are. Please add some more discussion to more accurately represent ESMs.

This sentence was deleted: “Each model has a specific way of representing the physical world.”

Pg 8, line 95. Need to clarify that SOM-FFN is just one realization, not the forced signal. Of course, it is all that we have, so the comparison to the multimodel mean is reasonable. The authors just need to make sure that the text here helps the reader to understand that observations are not the forced component.

We clarified this in the manuscript: “Unlike in ESMs, the observation-based product only represents the one realization of the real world, which includes internal variation, and is therefore not directly equivalent to the forced signal. However, the comparison to the 20 year mean multi-model mean still informs us about the degree of agreement between the two products.”

Table 1. Note in caption that internal is from CanESM only

Done!

Both Tables are difficult to read. Labels in column 2 are too small. The distinction between the scenarios is not clear enough.

Fair enough. This was not the format of the tables originally. We had to be change it to match submission requirements. The tables will be updated to match all your comments.

Figure 2. correct spelling in words in 2b

Will do! Thanks!

pg 11, line 65-66, “test from Santer et al. (2018)” should be defined in methods

A new section has been added to the Supplement.

pg 11, line 69. Strike “in the models” and replace with “in CanESM”

Done!

pg 12, line 77. Figure 2 (and also Figure 4) makes it evident that “model uncertainty” is much the mean spread across the models. Please mention this connection explicitly.

We could not quite understand what your comment refers to. There might be a typo in the comment here.

Pg 17, line 80-83. This section is poorly worded. Please rephrase to avoid “are mostly within hotspots but are not confined to them and do not include all of them”.. and to be more specific about the regions to which you refer.

The phrase “The regions of high internal variability (eastern boundary upwelling regions, western boundary currents of the Gulf Stream and Kuroshio, their extensions, and the Southern Ocean) are mostly within hotspots but are not confined to them and do not include all of them. This lack of correspondence explains why the correlation coefficients are not high for internal variability. “ was replaced with:

“Internal variability from CanESM5 is most dominant in mid-latitude eastern boundary upwelling regions and their extensions, in the North Atlantic, in the western boundary currents of the Gulf Stream and Kuroshio and their extensions, and in the Southern Ocean (Fig. 5). There is wide agreement between different models and estimation methods in regions of largest internal variability (comparison Fig. S4 added to the supplements). The regions of large internal variability are correlated with the same highly active regions for the sink anomalies discussed earlier (Fig 1c). However, correlation coefficients between internal variability and historical uptake are lower than those seen for scenario and model uncertainty.”

“Highly active regions” are also described more specifically in the supplement section S3.

Pg 17, line 87-93. This is just the mean sink, i.e. where low anthropogenic carbon is being brought to the surface. These regions continue to be ventilated from the deep and this is why the sink persists. There is no need to invoke teleconnections.

The final sentence on atmospheric telecommunications has been deleted and the paragraph ends like this:

“The model uncertainty is largest in the Southern Ocean consistent with CMIP5 models (Frölicher et al., 2015). Here, mode and intermediate waters are formed, and the complex nature of the sink varies on all time scales (Gruber et al. 2019). Frölicher et al. (2015) note the largest disagreement in ocean carbon uptake between models is in the Southern Ocean because the exact processes governing heat and carbon uptake remain poorly understood. The importance of model uncertainty in the Southern Ocean provides a clear focal point for modelling centers to concentrate their efforts in reducing projection uncertainty. “

Pg 17, Line 95-96. This description of the scenarios is not sufficiently precise. Scenarios are designed primarily to represent potential futures that socio-economic modeling indicates have potential to be realized. Within this range, there is a selection made of a representative pathways that are not too similar. But this is not the same as to say that they are "designed to deviate". See Riahi et al. 2017 http://dx.doi.org/10.1016/j.gloenvcha.2016.05.009

The sentence is replaced with:

"This is by construction as the scenarios deviate from each other with time to represent a range of pathways for future socio-economic possibilities in order to assess the long-term impacts of short-term decisions (Riahi et al., 2017)."

pg 18 line 01-03. please discuss what are these processes.

The discussion was elaborated on as follows:

"This shows that with pCO_2 differences across the air-sea interface interface being the main driver of the sink (Fay & McKinley, 2013; Landschützer et al., 2015; Lovenduski et al., 2007; McKinley et al, 2020; McKinley et al., 2017), the sink in these active regions evolves as the atmospheric CO_2 concentration changes because ocean processes associated with surface-depth connectivity constantly keep the surface ocean pCO_2 out of equilibrium with the atmosphere. In other words, the surface water in these regions are constantly renewed, mostly through advection and water mass formation, with water masses whose pCO_2 has not increased at the same rate as the atmosphere. Elsewhere, these conditions do not hold true and water at the surface equilibrates with the atmosphere on shorter time scales, decreasing the sensitivity to the projection scenario. "

pg 19, line 26 "fixed inactive regions"? please clarify. These regions are not "fixed" or "inactive"

"The fixed inactive regions, such as the centers of the mid-latitude gyre systems and the NE Pacific, show late emergence times and, in some cases, no detectability of the signal in any of the scenarios by 2100." was replaced with:

"Ocean regions such as the centres of the mid-latitude gyre systems and the NE Pacific show late emergence times and, in some cases, no detectability of the signal in any of the scenarios by 2100. Convergent large-scale circulation and strong stratification in these regions isolates the surface from the deep ocean reducing their capacity to hold large amounts of carbon (McKinley et al., 2016). An absence of mechanisms constantly drawing surface ocean CO_2 out of equilibrium with atmospheric CO_2 lets the surface water equilibrate with and adjust to the atmosphere on short time scales. Significant changes thus do not take place in the sink as the atmospheric CO_2 levels change and scenario uncertainty is lowest in the same regions (see Fig. 4). "

pg 20, line 41-51. The size of the forced trend is critical in the time of emergence. The scenarios with smaller forced trend emerge later. Please include this in the discussion

Thanks for the suggestion. Included!

Pg 21, line 72. Strike “basins”, replace with “regions”

Done!

Supplementary

Below Eqn S4. “Section S2”, instead of “Appendix B”

Edited!

Page numbers are needed in the Supplementary

Added.

Citation: https://doi.org/10.5194/esd-2022-19-RC1