Dear colleagues,

your revised manuscript went through review round yielding very conflicting responses. I therefore searched for a third, independent assessment, which took extra time. Based on this additional feedback, I am happy to inform you that your manuscript can now be published subject to minor revisions with editor review only.

We thank you and appreciate your detailed review of the manuscript which helps us make this a better contribution to the community. Here, we explain how we have addressed the reviewers' comments.

Please address the specific comments of reviewer #2:

"The authors have added references and more mechanistic description to the paper, as requested. This emphasizes the lack of novelty in the work. I have also taken the time to review the original review of Reviewer 2, and I concur with their finding that this paper is not providing new insights. The paper repeats analyses already presented, with the only distinction being that CMIP5 or a large ensembles was analyzed previously while here CMIP6 is analyzed. The update of model versions would not be expected to substantially change the behavior of the ocean carbon sink, and the authors show that this expectation is fulfilled. The specific calculation of "70% change in 40% of the area" is new, but is not particularly interesting because it has been know that some regions are intense carbon sinks has long been known.

We thank the reviewer for their time reviewing the revised manuscript. Our work does replicate and build on previous studies, but it also contains novel aspects. We update previous results with the new CMIP6 dataset, confirming that previous CMIP5 studies still hold. We also introduce new methodologies, including our large ensemble approach, and novel analyses, including our representation of sources of uncertainty across spatial scales. Our work brings together different ideas and techniques, previously tested separately, and frames them into a single study that addresses the ocean carbon sink based on the most recent version of CMIP models using a wide range of models and scenarios. Moreover, it provides quantification on different ideas that were known collectively through a combination of previous studies (such as our diagnosing the highly active regions, and performing a scale dependence analysis over a continuum of scales) into a coherent storyline that connects different pieces to each other.

Minor Comments:

Line 39-40: over what timeframe does this sentence apply?
Updated to:

"Despite increasing atmospheric CO_2 concentrations, the air-sea CO_2 flux does not change much in the middle of the subtropical gyres over the century starting in 1990. "

Line 42-44: This is incorrect. By definition, the response to increasing atm CO2 is, by definition, the anthropogenic sink.
Thanks for pointing this out. We would like to bring your attention to the following lines from Hauck et al. (2020):

"Note that this definition of the ocean carbon sink SOCEAN in the GCB is different from the definition of the "anthropogenic CO2 sink" referred to as the change in ocean carbon content only due to the direct effect of increasing CO2 concentration in the atmosphere ($F_{ant,ss}+F_{ant,ns}$), often used in the observational ocean carbon cycle community (e.g., Gruber et al., 2019)."

We have clarified the wording here in the manuscript to emphasize the direct absorption response and the changes to the natural background fluxes:

"The response of the ocean carbon sink to increasing atmospheric CO_2 levels consists of a direct absorption response as well as climate change induced perturbations to the natural background carbon fluxes (Crisp et al. 2022, McKinley et al. 2020, Hauk et al., 2020, Gruber et al. 2019, Frolicher at al, 2015). "

Line 47: which "different estimates"?
Added "(Landschützer et al., 2016,Gruber et al., 2009, Takahashi et al., 2009)"

Line 254: This “metric” is percent change in sink per area? Please be explicit as to what this metric is. Please also explain why it is useful, and explain of what this is a “metric”. Most importantly, what supposed to be learned by comparing models using this “metric” or by tracking it over time?

The metric is a criterion to identify the highly active regions. We have added a brief description of the criterion to the main text and refer the reader to the supplement for further information:

“Here, we provide a new criterion for identifying these highly active regions based on comparing the integrated global sink anomaly within grid cells above a certain threshold to the percentage of ocean area they occupy (see Supplement S5).”

In summary, this metric is a threshold that can *diagnose* the highly active grid cells from the rest of ocean grid cells, without having to make assumptions about the boundaries of the active regions for the sink. The resulting regions from this analysis match the regions of intense uptake change (trends) described in previous studies. To clarify its usefulness, we have added to the supplement: “The evolution of the threshold value with time, and the corresponding boundaries of the highly active regions, have implications for the evolution and efficiency of ocean carbon sink under the changing climate.”

Line 398-399: The word “active” is used twice. Please don’t use the word “active” to explain “active”.

In the mentioned lines, the word active is used twice to emphasize the reason for the term “highly active regions” - they actively respond to increasing atmospheric CO_2.

Figure 4. Please provide a figure title that immediately identifies this as NE Pacific and NW Atlantic . Please also mark NE Pacific and NW Atlantic on figures themselves.

Figure 4 shows essentially the same information as Figure 5. I don’t think Figure 4 is needed.

Thanks for pointing this out. We have updated the figure to label the regions in NE Pacific and NW Atlantic for each row of panels. We prefer to keep both Figure 4 and 5, because:

1- Figure 4 shows averages over specific regions in figure 5 which are not easily comparable visually on Figure 5;

2- Figure 5 alone does not reflect the effects of averaging on the signals as well as on the uncertainty;

3- Figure 4 provides intuition on comparing highly active and and not active regions as well as the association of scenario uncertainty with this designation using time-series and uncertainty ranges.
4- The removal of the figure would compromise the storyline.

Line 545. McKinley et al did not study CMIP5. They studied CESM-LE. Please correct Throughout

Thanks for pointing that out. They use both an ensemble of CMIP5 models and CESM-LE under the RCPs defined through CMIP5. We have clarified this throughout the paper.

And also the comments of reviewer #3:

The manuscript presents an analysis of uncertainties in ocean carbon uptake using methods that have previously been applied elsewhere to similar questions. The authors have nevertheless presented sufficiently interesting results that this work should be publishable as an incremental advance. However, I do believe that there are several points where there are misinterpretations of the model results, and misunderstanding of the connection to broader community efforts, and I believe that these critical (but minor by standards of revisions) issues should be clarified and regularized before the manuscript can be published.

We are thankful for the reviewer's constructive comments and welcoming review. We have addressed their specific concerns and comments as follows.

Overall there are two general arcs to the study, one geared towards quantitative analysis of uncertainty, and one geared towards interpretations and recommendations based on the analysis. It is the latter (the implications and interpretations) that are in particular need of attention. On Page 8 (lines 4-8) the authors state that the "largest change takes place in regions such as the North Atlantic, ….These regions of largest changes in the carbon sink seem to be the regions where there is a surface-depth connectivity. We refer to these regions as hotspots".

Whereas I fully agree with the authors that there are subduction hotspots where there is unambiguously a surface-interior exchange of properties via subduction, this in no way provides justification for arguing that gas exchange is dominated by these regions. As has been demonstrated by Iuidcione et al. (2016; Sci. Rep.) using GLODAP data, approximately 60% of anthropogenic carbon enters the ocean over the broad expanses of the subtropical cells, rather than in the hotspot regions highlighted here. This was also shown in the study of Rodgers et al. (JClim, 2020) with a CMIP5-class ocean model, where 60% of anthropogenic carbon uptake occurs over 45S-45N. Certainly there is enhanced uptake over western boundary current regions etc., but when uptake is analyzed in terms of density (Iudicone et al., 2016), models indicate that this enhancement of CO2 fluxes in western boundary currents is not equivalent to dominance in uptake.

As a related point, the authors state in the second paragraph of the Introduction (page 2, lines 36-38) that the air-sea flux doesn't change much over the subtropical gyres, with strong uptake being confined to regions with strong subduction. Again, in models where this has been evaluated (Iudicone et al., 2016) this is not true, so if the authors wish to make this point the burden is on them to demonstrate where previous studies are wrong.

Thanks for pointing out these interesting studies. We removed the term "hotspot" in our previous revision to avoid confusion, changing this to "highly active region".

More importantly, our discussion is not inconsistent with these previous studies. Ludicone et al. (2016) and Rodgers et al. (2020) describe values over the entire region of tropics and subtropics (including the western boundary currents), while we discuss more specific regional patterns. Importantly, there are large regions within the tropics and subtropics that are included in what we refer to as highly active regions (see Fig. S7). That said, Fig. 4 from Rodgers et al. (2020) is also showing that regionally, when compared to the pre-industrial state, the biogeochemically coupled model do not detect large changes except that they are concentrated at some regions which consistently match the regions we are *diagnosing* and describing as highly active (also shown in McKinley et al; 2016, Froelicher et al; 2015, etc.). Of course, a variety of combinations of ocean grid cells (regions) can explain 60% of the total anthropogenic uptake. What we focus on here is to account for the smaller scale regional differences when quantifying the highly active regions by also considering the area covered. Additionally, our diagnosis of highly active regions refers to their evolution in uptake as atmospheric CO_2 concentration increases, not to the sink itself. In other words, we focus on where the sink is accelerating rather than its absolute magnitude.

Finally, we have reviewed the manuscript and made sure that the text is not ambiguous in distinguishing uptake anomalies from the uptake itself, and makes clear where patterns of regional changes are considered and when averaging is taking place. Additionally, the lines in the introduction that the reviewer specifically pointed to were clarified as follows:

"Despite increasing atmospheric CO_2 concentrations, projected air-sea CO_2 fluxes do not change much in the middle of the subtropical gyres over the decade starting in 1990. The regions where ocean carbon uptake notably increases are those with strong exchange between the surface and the deep ocean (Ridge and McKinley, 2021; Frölicher et al., 2015; McKinley et al., 2016). "

Most importantly, the authors use some of these untested inferences to then propose that future optimization of the ocean observing system should be focused on the regions with large natural variability in carbon uptake. This is problematic for a number of reasons. First and foremost, it is not based on a quantitive assessment of observing system design for carbon. To do this, I believe the authors would be obliged to conduct an observing system simulation experiment (OSSE) as part of their study Second, even if an OSSE were to be conducted and demonstrate this for the case of carbon (which I do not believe would be the outcome of an OSSE), the authors really should refer back to the justification given through GOOS and others involved in planning for an optimized ocean observing system for ocean biogeochemistry, and make sure to reference correctly how any recommendations here fit their broader goal, if the authors intend to make such recommendations.

My recommendation is that the authors consider scaling back on their recommendations, by at least carefully qualifying them, in terms of community priorities and the fact that an OSSE has not been considered here. I think that the rest of the scientific presentation could stand in that case, and satisfy the requirements for sufficient incremental advance to warrant publication.

We can see the reviewer's point here and appreciate their concern. We have addressed their specific suggestions and scaled back on our recommendations, by pointing out only what our study explicitly suggests, acknowledging that we have not conducted an OSSE and pointing out caveats instead of explicit recommendations. We went through the manuscript and made changes accordingly. These include but are not limited to:

In the abstract: the finishing lines were changed to:

"In agreement with CMIP5 studies, our results suggest that to for a better chance of early detection of changes in the ocean carbon sink, and to efficiently reduce uncertainty in future carbon uptake, highly active regions, including the Northwest Atlantic and the Southern Ocean, should receive additional focus for modeling and observational efforts."

In section 3.4: Second paragraph, lines 17-19 updated to:

"Our results argue that focusing observational records inside efforts on the highly active regions are likely sufficient in order to detect human influence on the ocean carbon sink in the coming years/decades (2030-2050) if not earlier."

In conclusions:

- In the first paragraph, the following line was <u>deleted</u>:

This result implies that known regions of high historical uptake, are the same regions to prioritize for observing the future evolution of the sink.

- In the fifth paragraph, lines 6-8 were updated to:

“On the other hand, consistent observations in regions such as the Equatorial Pacific, the Gulf Stream and Kuroshio and their extensions, and the Southern Ocean, are likely to detect the emergence of the forced signal out of internal variability earlier in time.“