The authors investigate the sources of uncertainties in projections of air-sea CO_2 fluxes in CMIP6 ESMs and also provide estimates on the time of emergence of the forced signal.

I appreciate the time and effort that the authors put into developing this manuscript. However, I cannot recommend the paper in its present form. Even though I appreciate that the authors tackle an important question, namely the relative role of scenario uncertainty, model uncertainty and internal variability, it remained unclear to me what new insights are gained here in how the future ocean carbon sink evolves. As it stands, it is mostly an update of previous literature and analysis, but this time with CMIP6 models. In addition, I also have some concerns in how internal variability is estimated. See all my comments below.

We appreciate and thank you for your detailed review of the paper and constructive comments. It is of great value to know how we can make this paper a better contribution to the literature by addressing your comments and concerns. Hope you find our responses and edits satisfying.

Major comments:

Embedding results into existing literature

The detailed breakdown of uncertainty in the scenario uncertainty, model uncertainty and internal variability in air-sea CO_2 flux projections has been done by many others (e.g., (Lovenduski et al., 2016; Schlunegger et al., 2019, 2020)). Here the authors use CMIP6 models instead of CMIP5 models as used in those previous analyses, but the main results are basically the same as for the CMIP5 models: scenario uncertainty dominates at the global scale, followed by model uncertainty and then internal variability; time of emergence is early in high latitudes and in the tropics. If the authors want to publish this paper, the MS needs to include a thorough discussion on how these new CMIP6 results differ from what we already know from CMIP5. Or how it supports those previous findings. For example, the three last paragraphs in the discussion do not contain any single reference. However as mentioned above, many studies already exist who tackle similar questions. I am fine if the purpose of the paper is to give an update with CMIP6 models, but if so, this needs to be clearly stated upfront.

We have worked to embed the results into the literature by citing relevant previous studies more broadly. The aim of this study was to provide a quantification of the changes in the marine carbon sink and its sources of uncertainty based on CMIP6 models with updates in the methodology used and the analysis done compared to previous CMIP5 studies. We make use of an ensemble of 13 models to better capture model uncertainty in the response to different forcing (scenarios) and we use three scenarios to represent a wide range of future possibilities including a strong mitigation scenario. We have clarified and made sure in the abstract, introduction, and throughout the paper, in what way our analysis differs from the existing literature. For instance, our method to correct for the internal variability included in the spread across one realization of many models in order to estimate model uncertainty without having to apply a filtering or needing a large ensemble for every model is a novel approach. Moreover, we provide a novel analysis of how the three sources of variability change across the full continuum of scales, and a new metric to show how we can quantify the highly active regions for the sink. Finally, we have edited the manuscript to clearly show how our analysis builds on and supports previous studies and where our methods are different by citing the literature more heavily especially in the conclusions and results sections, and updating the discussions section.

The authors also highlight in the abstract that the ocean carbon sink is concentrated in highly active regions. That has been shown by many studies already (e.g. (Sarmiento et al., 1998)). Again, what is the novelty here?

We have now cited the mentioned paper in the manuscript where we bring up the discussion of highly active regions to show how our results support the previous studies. However, as pointed out in the abstract, we found that more than 70% of the change in the sink relative to the pre-industrial era is concentrated in less than 40% of the global ocean. This quantification of the highly active regions is another novel contribution that we now explain more clearly by defining our classification of highly active regions in the revised manuscript and supplement. We have added a new section to the supplement explaining the analysis done to quantify these highly active regions through a new metric. This metric finds the ocean grid cells over which the largest percentage of the integrated global sink is concentrated in the smallest percentage of the global ocean surface. The resulting highly active regions look as follows:

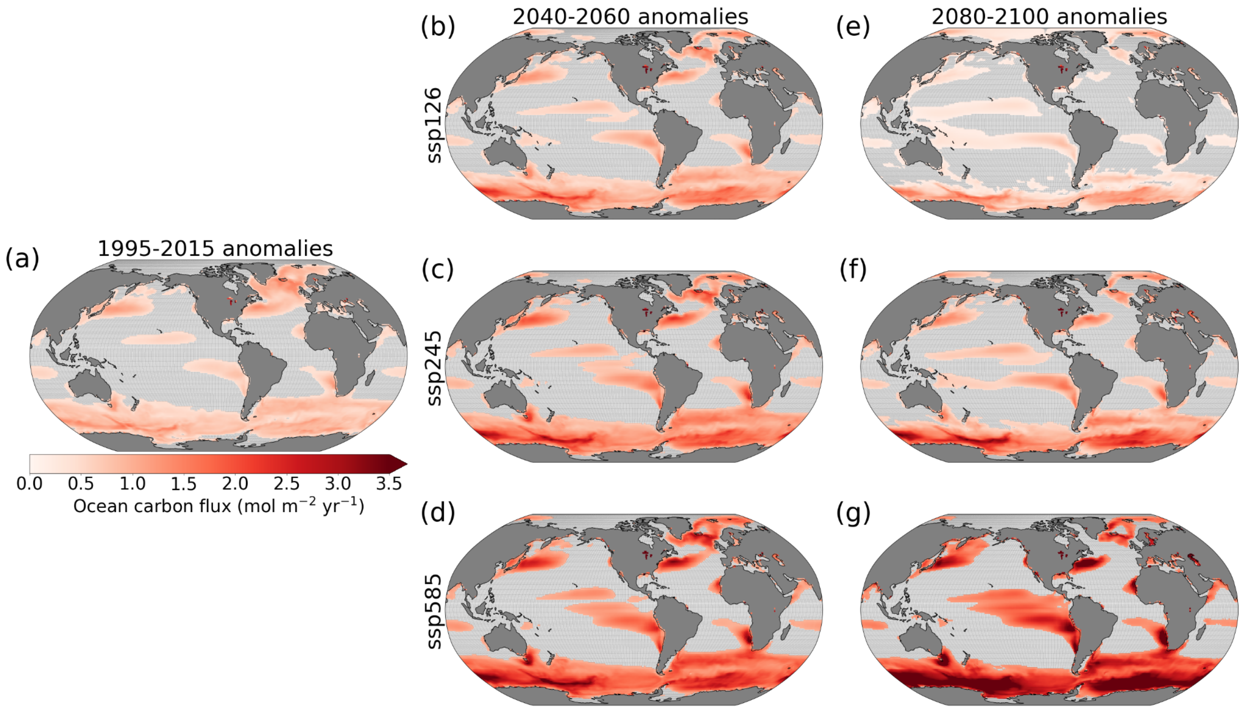

Figure S7- Highly active regions for the sink. The rows represent different scenarios and columns represent different time periods. The land is masked with grey color and grid cells outside of the "highly active regions" are hatched.

Calculating internal variability

The authors use one single ESM (i.e., CanESM5) initial condition large ensemble to estimate the internal variability in the air-sea CO2 flux. Whereas I see the benefit of using a large ensemble to estimate internal variability, as for example internal variability may be sensitive to changes in climate change and therefore changes with time, I suspect that the current results (fractional uncertainty and time of emergence) are heavily biased towards the CanESM5 model and how it represents internal variability. I suspect that different CMIP6 models simulate different magnitude of internal variabilities in air-sea CO2 fluxes. Therefore, the fractional uncertainty as well as the time of emergence might be different when using a different model. Therefore, it may make more sense to use piControl simulations from a variety of CMIP6 models (for example the same as used to estimate model uncertainty; if CMIP6 SMILEs are not available) to estimate internal variability and how uncertainties in the internal variability estimates impact the time of emergence and the uncertainty breakdown.

<u>(The following two questions were moved from your minor comments below to here:)</u>

L77- Data and methods: You assume that internal variability is well represented by CanESM5. But this might not be true.

L28-45- Data and methods: As explained above, I am not convinced to use one single model to estimate internal variability given that the models simulate a large spread in the magnitude of internal variability.

Thanks for pointing this out. We show in the supplements (Fig. S2) that, in order to use a large ensemble for the estimation of internal variability, the ensemble size must be large enough to sufficiently capture internal variation, and that internal variability has a scenario dependence that should be accounted for. In the ideal case, if every CMIP model provided sufficiently large SMILES for each scenario, an ensemble mean estimate of the variability could be obtained and would represent a best estimate (but still possibly biased compared to the real world). However, only a handful of CMIP6 models produced multiple ensemble members. We selected the CanESM5 SMILE as it is the only model that has a large enough ensemble over the entire timeline and set of experiments to make estimate internal variability robustly and across scenarios.

The use of a single model to estimate the scale of internal variability leads to some uncertainty in our estimates, as models do not agree perfectly with each other on the variability. Nonetheless, over the historical period, variability between large ensembles from three models that have a large enough number of ensemble members is within 10%, on the global scale (Fig S3). Moreover, the general patterns of the magnitude of internal variability are in good agreement across models and are consistent with known regions of high variability in the observed ocean (comparison Figure S4 is added to the supplements). Hence, the uncertainty imposed by using CanESM is very unlikely to influence our conclusions, although the quantitative details will vary. We have addressed possible biases that may arise from this choice throughout the revised manuscript.

In summary, the importance of internal variability decreases dramatically with time for the global ocean where internal variability is averaged out. Our analysis of fractional uncertainties is consistent with previous studies on both short and long time scales (Lovenduski et al.,2016, Schlunegger et al., 2020) on the global scale. Our goal for regional and local scale analysis is to point out the importance of the specific location in addition to the scale which is in agreement with previous studies, and the association between scenario uncertainty and how active the region is regarding the sink, which is valid and insensitive to the use of CanESM5. We agree that the results include characteristics specific to CanESM5 on regional scales. That could change the results for fractional uncertainties in the NE Pacific and NW Atlantic. Still, our tests on regional and local scales serve as an introduction for the need to understand patterns of the signal relative to internal variation and hence the TOE analysis for which our results are consistent with McKinley et al. (2016) and Schlunegger et al. (2020).

To address your concern and specific suggestion, we repeated the TOE analysis based on estimations of internal variability from the piContrtol runs of a subset of our CMIP6 multi-model ensemble for which the control runs were available. This leaves us with 11 models (excluded are CESM2 and NorESM2-LM). The results are presented below in comparison to the original analysis using CanESM5 LE.

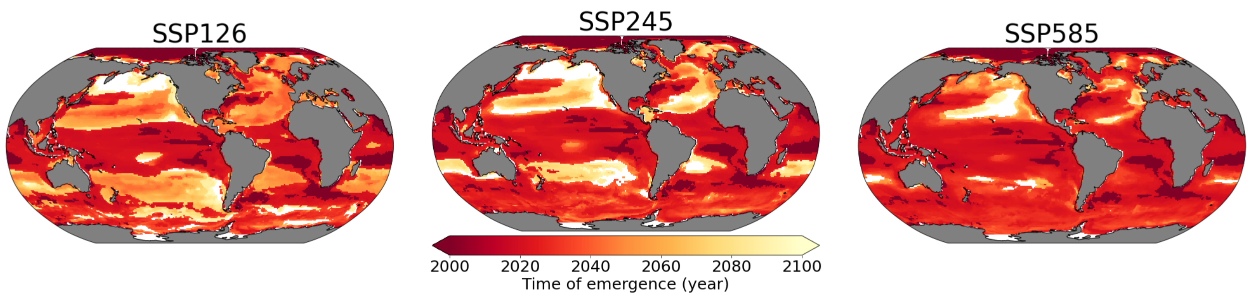

Fig a- TOE using internal variability estimation based on multi-model mean time variation of the century-long detrended piControl simulations (Frölicher et al, 2016).

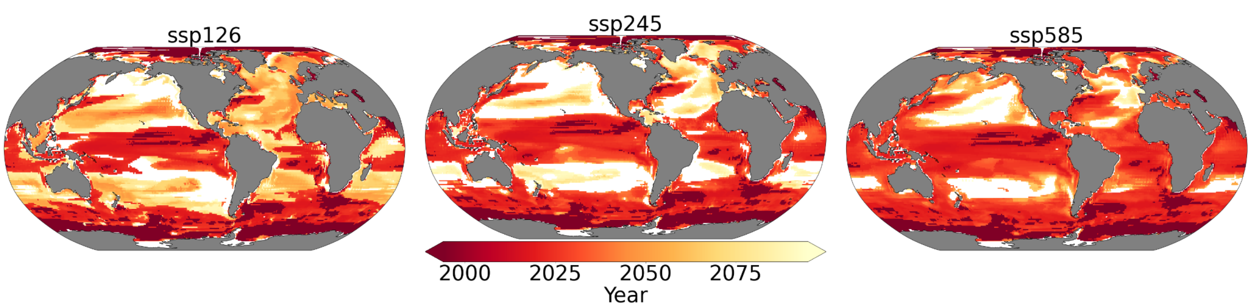

Fig b- TOE using internal variability estimation based on CanESM5 LE (Original version).

It is apparent from the figures that the broad patterns of earliest emergence occurring in the highly active regions (Equatorial upwelling regions, Southern Ocean, North Atlantic, western boundary flows and their extensions, etc.) and late/non emergence in the Ekman convergence regions in mid-latitudes are similar between the two. Moreover, the dependence on scenarios due to the strength of the forcing is also consistent between the two. The piControl simulations show earlier emergence and a smaller percentage of non-emergent grid cells on average, due to the smaller internal variability as a result of the stationarity assumption and independence of the forcing (scenario). All being said, the conclusions in the manuscript about the broad patterns and association with the discussed mechanisms are robust, while the magnitudes are somewhat different between the two analyses. This is most apparent in the Southern Ocean where TOEs are larger in piControl-based analysis which is attributed to the larger internal variability in the Southern Ocean in piControl-based simulations compared to CanESM5 LE (comparison Fig. S4 added to supplements).

In conclusion, both the CanESM5-based numbers and the piControl-based numbers are imperfect. However, the clear time/scenario dependence of internal variability and low error range between the available large ensembes makes our CanESM5-based analysis a more reliable estimate of the TOE. In response to your comment, we added this analysis comparing piControl-based TOE with CanESM5-based TOE, as well as a comparison map of different estimates of internal variability over the historical period to a new section in the supplements (section S3). Finally, we use the discussions of 1. the need to account for the change in internal variability with time and scenario, 2. the need for at least a certain number of realizations to capture internal variability in a large ensemble, and 3. the proposed methodology to correct for the internal variability included in model spread without needing to use any filtering (if we have a fair estimate of internal variability), to advocate for the need for more models to provide large ensembles so that the internal variability can be better estimated as the mean spread across multiple Large Ensembles.

Accounting for drift in air-sea CO2 fluxes in CMIP6 ESMs

Did you account for potential drifts in air-sea CO2 fluxes in the piControl simulations of the CMIP6 models? To estimate model uncertainty one should use the difference between the historical-scenario simulation and the piControl simulation (long-term trend). This would also allow to include the CNRM-ESM2-1 model as this model has a relatively large preindustrial outgassing of about -0.75 Pg C/yr, which is the reason why the ‘present-day’ CO2 flux is below all other models as shown in the Supplementary Figure S1. When correcting for this offset the model is close to all other models.

There are two different arguments here. One is the steady state flux and the other one is the drift. Since we use anomalies in our analysis, the steady state flux is removed. Hence, the correction does not explain why CNRM-ESM2-1 is an outlier. For a possible correction for the large outgassing (sum of different parameters such as model bias, outgassing due to riverine input, etc.), we need to make sure we do not remove a component of the sink that is crucial for comparison to the observation-based product. Given that, a deeper investigation of the model in the control simulations based on the CNRM-ESM2-1 model behavior assessment publications would possibly allow for the correction. However, the evaluation of the performance of individual models is beyond the scope of this paper.

We did not account for the drift and made sure that this is clearly stated in the revised manuscript. However, the drift is known to be small in the models compared to the historical trends (Hauck et al, 2020). Moreover, piControl simulations are not currently available for some of our selected models due to ESGF maintenance issues at some institutions. For 11 of our CMIP6 models for which piControl runs are available, on average, the drift is an order of magnitude smaller than the change in the model scenario with the smallest trend over the 21st century, on the global scale.

Accuracy of text

There are many places in the manuscript where the text is not accurate, and the reader might have difficulties understanding the details. For example, on page 11 l66, you state that ‘the trend in ocean carbon sink anomalies are statistically consistent between models and obs-based products based on tests from Santer et al. 2018. What test is this? The method has not been introduced in the method section.

We have included a new section in the supplement to explain the test. Thanks for pointing this out.

Minor comments (FYI: the line numbers are confusing in the MS)

Introduction:

L36-47: This paragraph mixes the description of the anthropogenic flux pattern with the total CO2 flux pattern. This is rather confusing.

The updated paragraph looks like this:

“The ocean’s capacity to absorb anthropogenic CO_2 is not uniformly distributed (McKinley et al., 2016, Sarmiento et al., 1998). Despite increasing atmospheric CO_2 concentrations, the air-sea CO_2 flux does not change much in the subtropical gyres. 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). This response of the ocean carbon sink to increasing atmospheric CO_2 levels consists of changes in both the anthropogenic and the natural carbon sink (Crisp et al. 2022, McKinley et al. 2020, Hauk et al., 2020, Gruber et al. 2019, Frolicher et al, 2015). Even within regions, there are large variations in the dominant mechanisms and the direction of the carbon sink. In the Southern Ocean for instance, the spatial superposition of natural and anthropogenic CO_2 fluxes leads to a relatively strong uptake band between approximately 55°S and 35°S (Gruber et al., 2019). However, south of the Polar Front (55°S), the different estimates agree less well (Gruber et al., 2019). Supported by measurements based on biogeochemical floats (Bushinsky et al., 2019; Gray et al., 2018; Williams et al., 2018), Gruber et al. (2019) argue that the region was most likely a small source in 2019. ”

L36: I guess there are many older papers that could be cited here (e.g. Takahashi or Sarmiento)

Thank you for the suggestions. We have added the Sarmiento (1998) paper to the citations, however, Takahashi does not talk about the anthropogenic sink specifically as far as our knowledge of the literature goes. Is there any specific paper you have in mind that we can refer to?

L38: Maybe cite here (Frölicher et al., 2015)

Agreed! Added to the citations.

l.52: Laufkötter et al. (2015) does not fit here as they do not look at air-sea CO2 fluxes. Maybe include (Terhaar et al., 2021) instead.

Thanks for the suggestion.

L54-59: Schlunegger et al. (2020) also investigated the sources of uncertainties in air-sea CO2 fluxes. This study should be mentioned here as well.

Edited as follows:

“Projection uncertainty varies with lead time, spatial averaging scale, and from region to region (Lovenduski et al., 2016; Schlunegger et al., 2020).”

L69: ‘along with others’: can you elaborate a bit what other processes you have in mind here? What about small-scale processes such as eddies, etc.?

The sentence was updated to:

“Modes such as the El Niño–Southern Oscillation, North Atlantic Oscillation, Atlantic Multidecadal Oscillation, Pacific Decadal Oscillation, and Southern Annular Mode (SAM) contribute to this internal variability. Internal variability also includes variability that acts on shorter time and spatial scales, such as submesoscale and mesoscale ocean features (Frolicher et al., 2016).”

l 71: ‘beyond short timescales’: Please backup this claim with a reference

The Following citations are added:

Somerville, R.C.J. The predictability of weather and climate. *Climatic Change* 11, 239–246 (1987). https://doi.org/10.1007/BF00138802

Lorenz E. N. The predictability of a flow which possesses many scales of motion. *Tellus.* 1969;21:19

l84: not only the physical world but also the biogeochemistry for example.

This sentence was also commented by our other referee and for the sake of accuracy, it was omitted from the manuscript.

L98-99: Can you update these number with CMIP6 estimates?

Wil be updated.

L99-00 (page 4): The introduction lacks a paragraph on earlier results. For example, Lovenduski et al. (2016), McKinley et al. (2016), and Schlunegger et al. (2020) have already tackled similar questions using CMIP5-type models and SMILEs. This needs to be stated upfront here.

Lovenduski et al. (2016), and McKinley et al. (2016) were cited in the introduction to review the literature. We have added a new paragraph advocating for the need to understand the patterns in TOE and have pointed to these earlier studies again, adding Schlunegger et al. (2020) that was not previously mentioned. Moreover, we have made sure to point out what approach was used in these studies and in what way our methods differ. The paragraph looks as follows (page 4, after the discussion about scenario uncertainty before the road map):

“Together with the patterns of changes in the sink, the patterns of internal variability allow for an assessment of the required timescales for detection of changes in the ocean carbon sink. Detection means that we can robustly separate the forced signal from internal variability (McKinley et al., 2016). Detectability can be assessed using Time of Emergence (TOE; Hawkins and Sutton, 2012; Lovenduski et al., 2016; McKinley et al., 2016; Rodgers et al., 2015; Schlunegger et al., 2020 & 2019). For example, McKinley et al. (2016) and Schlunegger et al. (2019) showed that the forced signal of increasing ocean carbon uptake is not detectable in the Ekman convergence regions of the subtropical gyres. Schlunegger et al. (2020) builds on that using four large ensembles of CMIP5 ESM simulations with two forcing scenarios to show that air-sea CO_2 flux TOEs show strong agreement between the large-ensembles not just for global and regional scales but also locally and spatially. Their use of only four models and two scenarios, however, potentially underestimates the contribution of model and scenario uncertainty.”

Data and Methods:

L07-15: Did you use CO2 concentration driven simulations or CO2 emission driven simulations. I guess the former, but please clarify.

Edited! We are using concentration driven simulations and this is clearly mentioned in the revised manuscript.

L54: I am a bit confused here. Do you correct here each model with the internal variability from the CanESM5? But what if the different models have a different internal variability than the CanESM5?

note: Variance* refers to the Mathematical function with the same name.

Equation S3 shows that if we decompose each models' one realization signal to the sum of the forced signal and a deviation due to internal variability, then by taking the Variance* across the multimodal first realizations (assuming that Covariance between "deviations due to internal variability" and "Forced signal" equals zero), the Variance* of the forced signal across many models equals the Variance* of the first realizations across the multimodal ensemble minus the Variance* of the "deviations from the forced signal due to internal variability" across the same models. Since internal variability is assumed to be random noise, in a large sample, the Variance* of the "deviations from the forced signal due to internal variability" across many models equals the internal variability. So, we can correct for internal variability included in the spread across different models' realizations if we have an estimate of internal variability.

The use of the CanESM5 LE as an estimate of internal variability is an assumption here clearly stated in the manuscript. As discussed in the manuscript and in detail earlier in this document, internal variability from CanESM5 LE is a fair approximation for internal variability based on current model data availability to account for nonstationarity in time based on scenario. Finally, we have revised the manuscript to make this correction more comprehensible. Thanks for the comment.

Results

L20: Maybe state in the first sentence of the Figure caption what quantity Figure 1 shows.

Edited!

Figure 2b: uncertainty is wrongly written in the Figure – Typo.

Edited!

Figure 2b: y axis: Fraction of what? Please clarify.

Explained in the caption and the figure is updated.

L66-68: How did you test this? How did you conclude that SOM-FFN shows a larger multidecadal variability? This is unclear?

We included a section on how we compare the trends from SOM-FFN to the multi-model mean in the supplements. The analysis shows the trends are consistent between the two products. Given that, the SOM-FFN smoothed time series (using a 10-year running mean) goes outside of the range of internal variability based on CanESM5 LE, which is on top of the multi-model mean signal (forced trend). It can also be shown that the variance of the SOM-FFN time-series is higher than any individual realization from the models.

We have included in both the text and Fig. 2 caption that all time-series are plotted after applying a 10-year running mean filter.

L77-78: Isn't that obvious, given that scenario uncertainty is zero over the historical period?

Yes, it is. The point here however was that in the historical period model uncertainty is a larger source of uncertainty compared to internal variability, and in the future scenario uncertainty starts to grow larger and larger until it becomes the dominant source.

L92-93: Schlunegger et al. (2020) shows it for many more regions. See their Figure 9.

Schlunegger et al. (2020) talk about how the partitioning of uncertainty differs from region to region and does not mention the magnitude of uncertainty. Still, we have referenced this study in the paragraph where it was relevant. Thanks for the suggestion.

L77-78: 'highly significant finding'. This has been shown already in (Wang et al., 2016). They show that models that simulate a small ocean anthropogenic carbon uptake over the last decades also simulate a small uptake over the 21st century.

Wang et al. (2016) discuss how the projected air-sea CO_2 fluxes are strongly associated with the simulated air-sea CO_2 fluxes in the historical condition, confirming that the historical state is a good predictor for the future state. We have cited their results in the revised manuscript to better back up our findings. However, throughout the paper, when we are talking about correlations, we are measuring spatial pattern correlations in contrast to Wang et al. (2016) estimation of magnitude correlations. Moreover, the phrase "highly significant finding" that you pointed out is referring to the pattern correlations between projected model/scenario uncertainty and the modern day flux anomalies, not the flux anomalies themselves. This is to mention that not only from Wang et al. (2016) we know that the historical state is a good predictor for the future state, we know it also informs us about the regions (patterns) of future uptake and model/scenario uncertainty.

L89-91: The large uncertainty in simulated uptake of Cant in the Southern Ocean simulated by ESMs has already been highlighted in previous studies (e.g. Frölicher et al. 2015)

We have made sure to cite the relevant literature with which the results from CMIP6 in this study are consistent. The edited section looks 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. ”

L41: Which previous studies? Please clarify.

The statement “Previous studies have largely presented a single time of emergence; however, the time of emergence strongly depends on the future scenario.” is updated to:

“Time of emergence strongly depends on the future scenario. Schlunegger et al. (2020) show for two scenarios that modest (~10 yr) TOE differences between different ESMs under strong anthropogenic forcing can evolve into pronounced (60+ yr) TOE differences with moderate mitigation. Here, we make use of three scenarios including a strong-mitigation scenario and account for scenario dependence of internal variability in our approximation using CanESM5. On average, scenarios with smaller forced trends emerge later as the size of the forced trend is critical to the time of emergence (Fig. 2-a).”

Discussions:

L63-64: Where is this shown.?There is no formal analysis on that in the paper.

Explanation added to the methods.

L69-97: All three paragraphs lack of any reference, even though many studies have investigated similar questions in the past. This needs to be changed.

The whole section is updated in this way:

“Ocean uptake of the increasing atmospheric CO_2 in the 21st century is concentrated in a few active regions with 70 percent of the total sink occurring in less than 40 percent of the global ocean. We analyze the results from the CMIP6 multi-model mean for the current state of the ocean (1995-2015), and the middle (2040-2060) and late (2080-2100) 21st century relative to the current state for three scenarios. We show that future changes in the sink are projected to mostly take place within the same historical highly active regions. This result implies that known regions of high historical uptake, including the North Atlantic and Southern Ocean, are the same regions to prioritize for observing the future evolution of the sink. Our results extend the argument of Wang et al. (2016) that the historical state is a good predictor of the future state to spatial patterns of change.

We show that the CMIP6 multi-model mean provides a consistent estimate of the spatial patterns of the sink, and the trend in the sink (globally), compared to the observation-based data product dataset of Landschützer et al. (2016). These results suggest the CMIP6 models are valid tools for understanding the past and future evolution of the ocean carbon sink, particularly at broad spatial scales. A notable area of disagreement is that the Landschützer et al. (2016) data shows larger decadal variability at the global scale than seen in any CMIP6 model and the range of internal variability from CanESM5 large ensemble. Gloege et al. (2021) shows that the SOM-FFN method overestimates the magnitude of decadal variability on the global scale due to the amount of gap filling.

We have shown that the magnitude of uncertainty and its partitioning among different sources differs with scale and location. On the global scale, scenario uncertainty is the largest source of uncertainty followed by model uncertainty and internal variability for CMIP6 models. These results are in agreement with previous studies form CMIP5 models (Lovenduski et al., 2016; Schlunegger et al., 2020). As the scales of integration (averaging) get finer, model and internal variability become the dominant sources, respectively. Testing the results on two ocean regions of about the same size, one in the NE Pacific and one in the NW Atlantic shows that - while consistent with the results of the scale dependence analysis - the relative importance of the sources of uncertainty also differs with location. Our test here extends the analysis Schlunegger et al. (2020) with a focus on the association of the location dependence with whether the regions have highly active carbon sinks. Notably, in highly active regions, such as the NW Atlantic, scenario uncertainty is large, whereas in more quiescent regions, such as the NE Pacific, internal variability is more significant. The dependence of internal variability on the scenario with time is another interesting finding that could be the subject of future studies for a better understanding of the driving mechanism and the degree of dependence on the future emissions and/or concentrations.

The patterns of high future CO_2 uptake uncertainty are highly correlated with the patterns of historical uptake. The correlation coefficients are highest for scenario uncertainty, indicating that the highly active regions have the potential for the sink to evolve according to the atmospheric CO_2 concentration, while the rest of the ocean basins do not respond strongly to changes in atmospheric CO_2 represented by the different scenarios. This finding has implications for assessment of the mitigations and effects of socioeconomic decisions, which is of relevance to the neighboring countries. Our results here are significant in that they show that regions of future uncertainty are largely associated with known regions of significant historical uptake.

Patterns seen in the time-of-emergence have implications for planning observational campaigns for detection of a signal (Schlunegger et al. 2019 & 2020). Furthermore, there is reverse association between how sensitive a region is to scenario differences (apparent in the scenario uncertainty patterns) and how sensitive the TOE is to scenarios. Our results show that caution should be taken in interpreting the observed changes in regions such as NE Pacific associated with late time of emergence of the signal from the decadal (internal) variations. On the other hand, regions such as the Equatorial Pacific, the Gulf Stream and Kuroshio and their extensions, and the Southern Ocean, should be the focus of consistent and expanded sampling for detection of the forced signal. Additionally, the patterns in sources of uncertainty show that model uncertainty is largest in the Southern Ocean, consistent with Frölicher et al., 2015. The sink in the Southern Ocean is driven by complex mechanisms involving coupled ocean-atmosphere-ice interactions that require better representation in ocean biogeochemical models. Significant progress in reducing uncertainties can be expected from new methods of bringing together models and observations (Frolicher et al. 2016). Our results provide a motivation to focus modelling as well as observational efforts on the known highly active regions of historical uptake.

Finally, we have shown that internal variability shows clear changes in time and depends on the scenario. The emergence of Large Ensembles (LEs) allows for quantification of these variations if enough ensemble members are available to fully capture internal variability using realizations that start from different initial conditions. Our use of the CanESM5 LE allows for us to account for the nonstationary of internal variability in time, like in Schlunegger et al. (2020) but with the advantage of also accounting for scenario dependence. Model intercomparison indicates that ESMs show differences in natural variability (Schlunegger et al. 2020). Nonetheless, our analysis of the global scale, of scale dependence, and of the patterns seen in Time of Emergence are consistent with previous studies, despite the potential sensitivity to the use of CanESM5 LE. Our methodology to correct for internal variability from model spread, without filtering or having a large ensemble for each ESM (which would limit the number of ESMs that can be included and, consequently, underestimate model uncertainty) lays the foundation for future studies when LEs are available from more ESMs and advocates for more modelling groups to provide such LEs in order to achieve an even more robust estimate of internal variability as the mean across different ESMs.”