The authors investigate the sources of uncertainties in projections of air-sea CO2 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.

Major comments:

Embedding results into existing literature

The detailed breakdown of uncertainty in the scenario uncertainty, model uncertainty and internal variability in air-sea CO2 flux projections has been done by many others (e.g.,
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.

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?

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.

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.

Accuracy of text

There are many places in the manuscript where the text is not accurate, and the reader might have difficulties to understand the details. For example, on page 11 l66, you state that ‘the trend sin 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.

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.

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

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

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.
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.

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.?

L71: ‘beyond short timescales’: Please backup this claim with a reference

L84: not only the physical world but also the biogeochemistry for example.

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

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.

Data and Methods:

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

L28-45: 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.

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?

L77: You assume that internal variability is well represented by CanESM5. But this might not be true.
Results

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

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

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

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

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

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

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.

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)

L41: Which previous studies? Please clarify.
Discussions:

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

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.

References


Schlunegger, S., Rodgers, K. B., Sarmiento, J. L., Ilyina, T., Dunne, J. P., Takano, Y.,
