I am deeply thankful to the anonymous reviewer for their constructive and very helpful evaluation of the manuscript, which has strongly improved its quality. Below, I have addressed the points that were raised by the reviewer point by point. The reviewer's text is shown in black and my responses in blue. Text that will be part of the revised manuscript is shown in italic and blue color.

This manuscript is a nice addition to the long-standing scientific debate about what drives decadal to multi-decadal variability in the ocean carbon sink. The manuscript is comprehensive and mostly well written and clearly structured. The methods and results are well explained and appear robust.

My one truly major comment is regarding the lack of proper statistical analysis. This will have to be added before the manuscript is published. In addition I want to raise several other issues that should be considered and which I think would help improve the manuscript and increase its impact.

Response: Thank you for your positive and constructive evaluation of the manuscript.

General:

Occasionally the reading is marred by awkward sentences. The second sentence in the abstract exemplifies this: Observing the ocean carbon sink is not challenging because there are high uncertainties, but the uncertainties are high because observing the ocean is challenging.

Response: Thank you for your advice to improve the English. The sentence was adjusted as suggested:

"Despite the ocean's importance for the carbon cycle and hence the climate, uncertainties of the decadal variability of this carbon sink and the underlying drivers of this decadal variability remain large because observing the ocean carbon sink and detecting anthropogenic changes over time remain challenging."

Throughout the manuscript the author states that observation-based pCO2 products extrapolate observations. I do not like the use of the word "extrapolate" here. When we extrapolate we extend into an unknown situation, but the observation-based products primarily attempt to interpolate between known situations. It is a bit nit-picky, and probably boils down to semantics, but I'd prefer the term "gap-filling". In my mind that is more comprehensive.

Response: I slightly prefer to keep the word "extrapolation" to guarantee consistency with previous literature. Fay et al. (2021) provided a reference for other studies when they created a harmonization 6 such pCO_2 products with their developers as co-authors. In this study, they use the word 'extrapolation'. However, I have no strong opinion and am happy to change 'extrapolate' to 'gap-filling' if this is preferred by the reviewer and the editor.

Introduction:

The introduction begins with a statement that the ocean has removed about 25% of all anthropogenic emissions since the onset of the industrial revolution. This is not factually incorrect, but I still find the statement somewhat misleading. Based on Table 8 in Friedlingstein et al. (2023) the cumulative ocean uptake (both since 1750 and since 1850) amounts to approximately 25% of the total emissions (FF+LUC). However, given this statement at the beginning of section 3.9 in the same paper "The cumulative land sink is almost equal to the cumulative land-use emissions (220±70 Gt C), making the global land nearly neutral over the whole 1850–2022 period." we understand that ocean has been more important in storing human-made emissions than the "has taken up around one quarter of all anthropogenic emissions" would indicate. In a paper highlighting the ocean sink I think this is a nuance worth noting in the introduction. But I stress again that the statement is not actually incorrect.

Response: As suggested by the reviewer, a sentence was added to the Introduction:

"If land-use change emissions are considered part of the land carbon sink, the land becomes almost neutral and the ocean carbon sink becomes the only major natural carbon sink (Friedlingstein et al., 2023)".

In the introduction (lines 92 onwards) the point is made that differences between observation-based products and GOBM products could be due to the data sparsity. Here it should be noted that it is not the sparsity that is the major problem, but rather the uneven sampling in time and space. Hauck et al. (2023) showed that if observations were regularly spaced the differences largely disappear. This is worth mentioning because it is a problem much more difficult to remedy than just having too few observational data.

Response: Following the suggestion by the reviewer, the sentence was changed to:

"As opposed to the magnitude, the differences in the decadal trends between pCO_2 products and GOBMs might be due to uneven sampling of observations in space and in time, e.g., few observations in the 1980s and 1990s and few observations in the southern hemisphere, as demonstrated with a subset of pCO_2 products evaluated with output from a GOBM (Hauck et al., 2023) and ESMs (Gloege et al., 2021)."

Throughout the manuscript it can be difficult to understand what "drivers of the decadal trends" actually mean. My understanding is that this study looks at defining the underlying causes of multi-decadal variability, that is, variability in the decadal trends. I could be mistaken, but this should regardless be stated more clearly in the introduction.

Response: Following the suggestion by the reviewer, the respective sentence in the Introduction was extended for clarification:

"Here, I use an ensemble of 12 ESMs from phase 6 of the Coupled Model Intercomparison Project (CMIP6) (Table 1) to provide a new perspective on potential drivers of the decadal trends of the ocean carbon sink, i.e., the underlying causes of its multi-decadal variability."

In the introduction several limitations to studies using the observation-based and GOBM products are presented. This is correct and fair, but there are also limitations to using ESMs so a few sentences describing these should be added at the end of the introduction.

Response: As suggested by the reviewer, we have added the following text about ESMs to the Introduction:

"Here, I use an ensemble of 12 ESMs to provide a new perspective on potential drivers of the decadal trends of the ocean carbon sink from phase 6 of the Coupled Model Intercomparison Project (CMIP6) (Table 1). Fully coupled ESMs are another tool to quantify and understand the ocean carbon sinks (e.g., Joos et al., 1999; McNeil and Matear, 2013; Frölicher et al., 2015; Goris et al., 2018; Terhaar et al., 2022b, 2021b). As ESMs are fully coupled and not forced with atmospheric reanalysis data, they do not simulate the same inter-annual internal climate variability as pCO_2 products and GOBMs do and their biases of the surface ocean physics and biogeochemistry are thus larger than surface ocean biases of GOBMs (Terhaar et al., 2022b; Terhaar et al., 2024). However, ESMs have distinctive advantages compared to pCO_2 products and GOBMs for the analyses of decadal drivers of the ocean carbon sink because (1) they cover a period of 251 years from 1850 to 2100, (2) cover at least four different future scenarios, and (3) they all have a different internal climate state."

Methods:

In section 2.2 the author describes how the magnitude of the ocean carbon sink in the 12 different ESMs was corrected/adjusted. First, I would think that the magnitude of the ocean carbon sink is related to the model's climate state since it is closely linked to the ocean circulation. So what are the consequences of doing such a correction to the models? Second, it is stated as fact that a "negative bias in the magnitude of the carbon sink also introduces a negative bias in the decadal trends". I do not understand why this will always be the case. This part of the method requires a bit more description, and a bit more discussion.

Response: The adjustment allows to bias-correct simulated output based on previously identified biases in the ocean circulation and surface ocean carbonate chemistry (Terhaar et al., 2022b).

Below, I have re-made the main figure from the manuscript without the adjustment. Without the bias-adjustment for each Earth System Model, the strength of the relationship slightly reduces from r^2=0.90 to r^2=0.83 but remains strong and significant. Hence, the findings are not sensitive to the bias-adjustment, but the bias-adjustment still yields more reliable results as the studies from Goris et al. (2018), Terhaar et al. (2022b) and Terhaar, Goris, et al. (2024) suggest. For clarification, I added the following sentence to the Methods:

"The adjustment corrects for known biases in the models' circulations and surface ocean carbonate chemistry and hence reduced differences in the overall magnitude of the simulated carbon sink between ESMs (Terhaar et al., 2022b). This reduction in the difference in the magnitude of the carbon sink also reduces differences between the magnitude of trends and slightly improves the relationships found here as it (r^2 in Figure 3 would have been 0.83 without

adjustment instead of 0.91 with adjustment). Nevertheless, the results would quantitatively and qualitatively almost identical with and without that adjustment. "

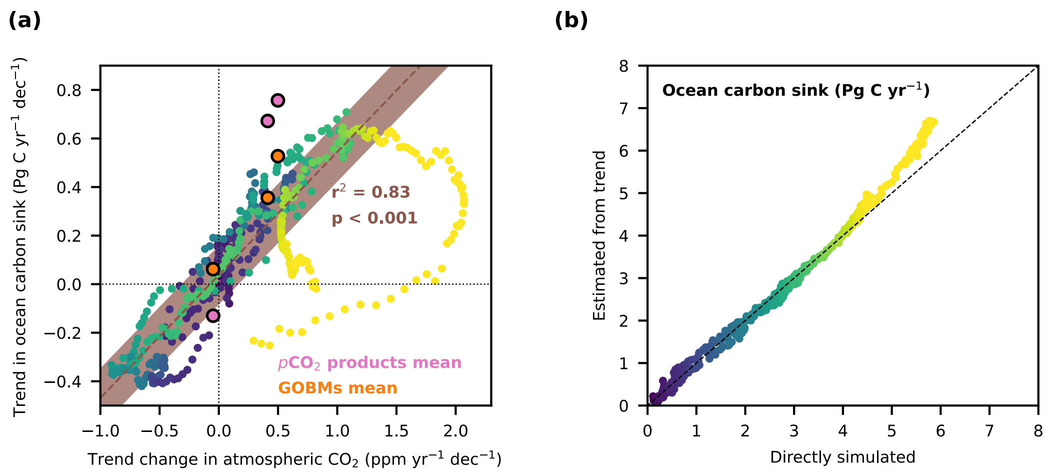

Figure 3: The relationship between changes in the atmospheric CO2 growth rate and decadal trends of the global ocean carbon sink for the multi-model mean. **(a)** Decadal trends of the multi-model mean ocean carbon sink compared to changes in decadal trends in atmospheric CO2, which represent the decadal averaged growth rate of atmospheric CO2. The dark blue to yellow circles without a surrounding black line show multi-model averages for all years of the historical period from 1850 to 2014 and for all years from 2015 to 2100 for all four SSPs. All decades over from 1850 to 2100 are shown, i.e., 2000-2009, 2001-2010, 2002-2011, etc. The brown line shows a linear fit for all years when the global ocean carbon sink is smaller than 4.5 Pg C yr-1 and the brown shading is the 1-s projection uncertainty. The dots with black lines around them show values from the respective ensemble means of the pCO2 products (pink) and GOBMs (orange) from the Global Carbon Budget 2023 (Friedlingstein et al., 2023) for the three decades between 1990 and 2020. **(b)** The simulated ocean carbon sink in comparison to the expected ocean carbon sink based on the relationship in **(a)** and the prescribed trend change in atmospheric CO2 in the simulations.

In the introduction the author argues that one of the benefits of using ESMs over other types of data products is the ability to perform robust statistics. I completely agree, but it seems as if no, or very little, statistical analysis has been performed. At least there is no description of such analyses in the methods. This I think is a major weakness of the manuscript as it stands now. At the very least a table with statistics for Figures 3, 4, 5 should be added, but a more thorough statistical approach to the analysis in section 5 and 7 would also be beneficial. I also note here that the author presents a p-value for the regression analysis (Figures 3, 5, 8). The p-value has been a topic of discussion for several years now, and is often misused. At the very least you must state what null hypothesis you are testing. However, given that you have a lot of data points the p-value is perhaps not the most useful metric. Given the amount of data available when using several ESMs I would suggest you try a Bayesian hypothesis testing instead for a more useful metric for significance.

Response: Thank you for your comment. While my argument in the Introduction was a comparison of the statistics that can be done over a few decades with only one climate mode, so really no statistics, and the statistics that can be done over many decades and different scenarios that are provided by ESMs, I now see that the p-value might not be enough. Having taken the time to learn more about the literature in statistics, I have now added the following section to the Methods:

"2.7 Coefficient of determination, p-values, and Bayes factor

To determine the strength of correlations, the coefficient of correlation was calculated throughout this study (r^2). In addition, the p-value was calculated to test the hypothesis that the trend change in atmospheric CO_2 from one decade to another decade is a significant driver of trends of the global and regional ocean carbon sink. A p-value larger than 0.1 indicates little or no evidence or that hypothesis exists, a p-value from 0.1 to 0.05 indicates weak evidence or a suggestion of evidence, a p-value from 0.05 to 0.01 indicates evidence or modest evidence and a p-value from 0.01 to 0.001 indicates strong evidence (Held and Ott, 2018). In addition, an upper bound for the Bayes factor can be calculated following Halsey (2019). Throughout the manuscript, the p-values are never larger than 1e-89 resulting in Bayes factors that are at least 1e86. Based on the Bayes factor the hypothesis that the trend change in atmospheric CO_2 from one decade to another decade is a significant driver of trends of the global and regional ocean carbon sink is 1e86 more likely than the hypothesis that the trend change in atmospheric CO_2 from one decade to another decade is not a significant driver of trends of the global and regional ocean carbon sink. As p-values are that small and Bayes factors are that high, I simply refrain to report that the p-values are smaller than 0.001."

As written in this new section on the manuscript, I have refrained from adding an additional table as the p-values are extremely small and the Bayes factors are extremely large, both indicating strong confidence in the identified relationship. In addition, the relationship is found on explained mechanisms based on McKinley et al. (2020), which further supports the argument that it is not a randomly emerging relationship.

Results:

In Figures 1 and 2 it is hard to tell which line represents which scenario. Please choose colors with more contrast.

Response: The colors were chosen in accordance with the IPCC report for the SSPs to be recognizable by most readers. Therefore, I prefer not to change the colors.

In Figures 3-5 it is difficult to tell the difference between the observation-based points and the GOBM-based points. They are both too small (given the black outline), and the colors are too similar to easily differentiate.

Response: When I chose the colors orange and pink, I tried to find colors that are easier to distinguish, that are different enough from the viridis colormap used for the ESMs and are also distinguishable for color blind people. During the revisions, I tested different other options but could not find a better combination of colors. I am happy to change colors if the reviewer or the editor finds a better solution.

In Fig. 3, the dots size was slightly increased. In Figure 5, increasing the dot size is unfortunately not possible as larger dots would hide other dots. The black outline is necessary to guarantee clear visible difference to the dots for the ESMs.

I find Figure 4 very interesting and would have liked more discussion about it. It looks like there may be some regional variation in at what global sink strength the regional sink deviates from the expected trend. Intuitively this makes sense to me, but it would be interesting to see whether it really is the case or just me seeing things. Either way it would add interesting discussion about why the relationship breaks down. Also, considering that the regional analysis, for which no correction of the low bias in sink magnitude was performed (section 2.2), produces results so comparable to the global analysis, why is the bias-correction in section 2.2 necessary? It would be good if it could be tested what the results would be if no correction was done.

Response: The test for the difference with and without correction was provided in the answer above. Indeed, the correction has only a small effect. Nevertheless, I still prefer to keep it as it is always better to correct for known biases, even if the effect is rather small. Regionally the r^2 are almost all the same as globally (when calculated without correction). The only difference is in the Arctic Ocean, which is already discussed in the manuscript:

"The correlation coefficient is larger than 0.84 in the Atlantic, Pacific, Indian, and Southern Ocean. Only the Arctic Ocean has a smaller correlation coefficient of 0.66.

In the Arctic Ocean, the carbon sink has been shown to be already substantially more affected by climate change than in any other ocean basin (Yasunaka et al., 2023). In the future, when sea ice will disappear and the Arctic will continue to warm faster than any other region, the importance of climate change for the Arctic Ocean carbon sink will likely remain relatively large, for example through freshening (Terhaar et al., 2021a) and a change in the seasonal cycle of pCO_2 (Orr et al., 2022), and hence reduce the importance of changes in the atmospheric CO_2 for trends in the ocean carbon sink."

In addition, the three brief episodes of a few years where the relationship does not hold in the Southern Ocean are also mentioned although the overall relationship is strongest in the Southern Ocean despite these three episodes. While these three episodes have no effect on the relationship, an analysis and explanation for this deviation would interesting. In the manuscript, I provide an explanation for two of the three periods, when the deviations are also globally visible, but I cannot provide an explanation for the period from 1995 to 2005:

"The time periods where the differences are visible globally (2030-2050 and 2080 to 2100 under SSP1-2.6) are the times when the growth in atmospheric CO_2 stops and when it starts to decrease in that scenario (Fig 1c). As the atmospheric CO_2 growth rate changes quickly in these periods (Fig. 2a), first by changing into a decreasing phase and then transitioning into a stabilizing phase, it appears that a fast transition of the trend change in atmospheric CO_2 temporarily leads to differences in the expected relationship. If the trend change in atmospheric

CO_2 decreases fast, the trend in ocean carbon sink remains larger than expected and if the trend change in atmospheric CO_2 increases fast, the trend in ocean carbon sink remains smaller than expected. However, the drivers behind the divergence from the expected decadal trend of the multi-model mean in from 1995 to 2005 in the Southern Ocean remain unclear and should be analysed in future research."

In addition to this assessment, an in-depth analysis of the response in the Southern Ocean for the time period from 1995 to 2005 would be needed, which, however, extends the scope of this manuscript. Hence, no further discussion is added.

Section 4 warrants a more robust statistical analysis and more discussion. Right now no reasons are given for the presented differences. Also, given the short timeframe for most of the observation-based products how robust are the presented results?

Response: The difference between the expected trends based on ESMs and $p\mathrm{CO_2}$ products will also need much more work, especially on the side of the $p\mathrm{CO_2}$ products. This study here focuses on the ESMs and what can be expected in trends. I believe that the ball is now on the side of the $p\mathrm{CO_2}$ products to explain why trends are that large in these products. Part of the explanation was given by Hauck et al. (2023) as discussed in the manuscript.

The timeframe should be no problem. As long as these $p\mathrm{CO_2}$ products and GOBMs cover at least one decade, that decade can be compared to the expected trend as the driver of that trend, the change in atmospheric CO_2 is known for much longer timeperiods.

Minor comments:

Line 26: add "and" before "the ocean"

Response: Changed as suggested.

Line 71: in the abstract you state "3 to 7 decades", here it is "four to seven"

Response: Changed to "three to seven".

Line 75: the observations are not just relatively sparse, they are very sparse

Response: The word "sparce" was removed following the reviewers' comment.

Line 88-89: this sentence is unclear and needs rewriting for clarity

Response: The sentence was modified and divided into two sentences for clarification.

Line 104: move "from phase 6 of the Coupled Model Intercomparison Project (CMIP6)" to directly after "12 ESMs" on line 103

Response: Changed as suggested.

Line 158: change “effect” to “affect”

Response: Changed as suggested.

Line 190: It is not easy to see these jumps in the figure. Consider highlighting them somehow (different colors?)

Response: The jumps are now marked in the figure as suggested. Please see response to major comment of reviewer 2 (Prof. Galen McKinley).

Line 240: replace the first “and” by “with”, and the second “and” by “or”

Response: Changed as suggested.

Line 261-262: This sentence is incomplete

Response: The sentence was changed to:

“Once the atmospheric CO_2 growth declines, the trend in the ocean carbon sink becomes negative.”

Line 305: It is unclear whether this is the Pearson correlation coefficient (r) or the coefficient of determination (r^2). Based on the rest of the section I would guess the latter, but please specify and use the correct terminology.

Response: Changed as suggested.

Line 322-323: This sentence (beginning with “As the atmospheric …”) is incomplete

Response: The sentences was corrected to:

“As the atmospheric CO_2 growth rate changes quickly in these periods (Fig. 2a), first by changing into a decreasing phase and then transitioning into a stabilizing phase, it appears that a fast transition of the trend change in atmospheric CO_2 temporarily leads to differences in the expected relationship.”

Line 351: earlier in the manuscript “we” is used – be consistent

Response: “We” was changed to “I” as suggested.

Line 378: add “final” before GOBM

Response: Changed as suggested.

Line 390: “causes changes”

Response: Changed as suggested.

Figure 6: Add details in caption about the vertical line and shading in subplot c)

Response: The details were added as suggested.

Figure 9: It is next to impossible to tell the lines on this figure apart. Please choose different colors. I would also recommend making the lines for individual models thinner.

Response: The thin lines were intended to show the range of the individual models. As the previous version of the figure has not succeeded in doing so, the thin lines were replaced by a lighter shading that indicates the maximum and minimum of the variability of the decadal trends in the ESMs. The revised figure looks as follows:

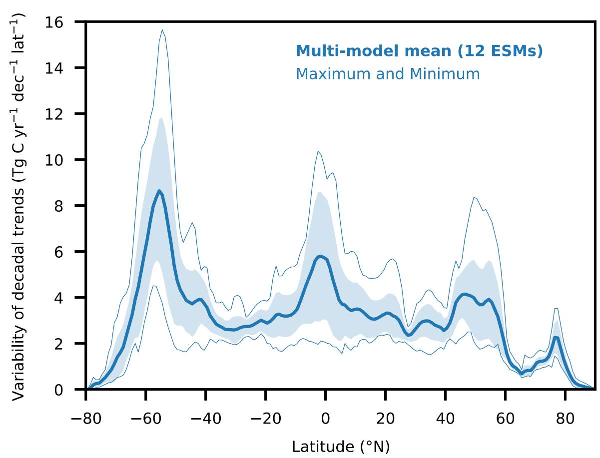

Figure 9: Variability of the decadal trends of the zonally integrated ocean carbon sink in earth system models. The multi-model mean (thick blue line) and the 1-σ standard deviation of the variability of the zonally integrated ocean carbon sink across the 251 years of the pre-industrial control simulation across all 12 ESMs. In addition, the maximum and minimum variability in the ESMs are shown at each latitude (thin blue lines).

Line 563: Are these the same five in every decade? Please specify and if not this warrants more discussion

Response: Following the suggestion of the reviewer, the following sentences were added to the manuscript:

"Only two pCO_2 products (NIES-ML3 from Zeng et al. (2022) and OS-ETHZ-Gracer from Gregor and Gruber (2021)) lie within the 1- σ and 2- σ ranges in the 1990s and 2000s, and only very slightly above the 2-σ range in the 2010s. The slightly higher trend in the 2010s in these products may very well be a consequence of the uneven sampling in space and time (Hauck et al., 2023). While the trends in these two pCO_2 products are closer to what is expected based on ESMs, only an in-depth analysis will eventually allow with to judge the performance of each pCO_2 product with certainty."

References

Arora, V. K., Katavouta, A., Williams, R. G., Jones, C. D., Brovkin, V., Friedlingstein, P., Schwinger, J., Bopp, L., Boucher, O., Cadule, P., Chamberlain, M. A., Christian, J. R., Delire, C., Fisher, R. A., Hajima, T., Ilyina, T., Joetzjer, E., Kawamiya, M., Koven, C. D., Krasting, J. P., Law, R. M., Lawrence, D. M., Lenton, A., Lindsay, K., Pongratz, J., Raddatz, T., Séférian, R., Tachiiri, K., Tjiputra, J. F., Wiltshire, A., Wu, T., and Ziehn, T.: Carbon–concentration and carbon–climate feedbacks in CMIP6 models and their comparison to CMIP5 models, Biogeosciences, 17, 4173–4222, https://doi.org/10.5194/bg-17-4173-2020, 2020.

Fay, A. R., Gregor, L., Landschützer, P., McKinley, G. A., Gruber, N., Gehlen, M., Iida, Y., Laruelle, G. G., Rödenbeck, C., Roobaert, A., and Zeng, J.: SeaFlux: harmonization of air–sea CO2 fluxes from surface pCO2 data products using a standardized approach, Earth Syst. Sci. Data, 13, 4693–4710, https://doi.org/10.5194/essd-13-4693-2021, 2021.

Friedlingstein, P., O'Sullivan, M., Jones, M. W., Andrew, R. M., Bakker, D. C. E., Hauck, J., Landschützer, P., Le Quéré, C., Luijkx, I. T., Peters, G. P., Peters, W., Pongratz, J., Schwingshackl, C., Sitch, S., Canadell, J. G., Ciais, P., Jackson, R. B., Alin, S. R., Anthoni, P., Barbero, L., Bates, N. R., Becker, M., Bellouin, N., Decharme, B., Bopp, L., Brasika, I. B. M., Cadule, P., Chamberlain, M. A., Chandra, N., Chau, T.-T.-T., Chevallier, F., Chini, L. P., Cronin, M., Dou, X., Enyo, K., Evans, W., Falk, S., Feely, R. A., Feng, L., Ford, D. J., Gasser, T., Ghattas, J., Gkritzalis, T., Grassi, G., Gregor, L., Gruber, N., Gürses, Ö., Harris, I., Hefner, M., Heinke, J., Houghton, R. A., Hurtt, G. C., Iida, Y., Ilyina, T., Jacobson, A. R., Jain, A., Jarn\'\iková, T., Jersild, A., Jiang, F., Jin, Z., Joos, F., Kato, E., Keeling, R. F., Kennedy, D., Klein Goldewijk, K., Knauer, J., Korsbakken, J. I., Körtzinger, A., Lan, X., Lefèvre, N., Li, H., Liu, J., Liu, Z., Ma, L., Marland, G., Mayot, N., McGuire, P. C., McKinley, G. A., Meyer, G., Morgan, E. J., Munro, D. R., Nakaoka, S.-I., Niwa, Y., O'Brien, K. M., Olsen, A., Omar, A. M., Ono, T., Paulsen, M., Pierrot, D., Pocock, K., Poulter, B., Powis, C. M., Rehder, G., Resplandy, L., Robertson, E., Rödenbeck, C., Rosan, T. M., Schwinger, J., Séférian, R., et al.: Global Carbon Budget 2023, Earth Syst. Sci. Data, 15, 5301–5369, https://doi.org/10.5194/essd-15-5301-2023, 2023.

Frölicher, T. L., Sarmiento, J. L., Paynter, D. J., Dunne, J. P., Krasting, J. P., and Winton, M.: Dominance of the Southern Ocean in Anthropogenic Carbon and Heat Uptake in CMIP5 Models, J. Clim., 28, 862–886, https://doi.org/10.1175/JCLI-D-14-00117.1, 2015.

Gloege, L., McKinley, G. A., Landschützer, P., Fay, A. R., Frölicher, T. L., Fyfe, J. C., Ilyina, T., Jones, S., Lovenduski, N. S., Rodgers, K. B., Schlunegger, S., and Takano, Y.: Quantifying Errors in Observationally Based Estimates of Ocean Carbon Sink Variability, Global Biogeochem. Cycles, 35, e2020GB006788, https://doi.org/https://doi.org/10.1029/2020GB006788, 2021.

Goris, N., Tjiputra, J. F., Olsen, A., Schwinger, J., Lauvset, S. K., and Jeansson, E.: Constraining Projection-Based Estimates of the Future North Atlantic Carbon Uptake, J. Clim., 31, 3959–3978, https://doi.org/10.1175/JCLI-D-17-0564.1, 2018.

Gregor, L. and Gruber, N.: OceanSODA-ETHZ: a global gridded data set of the surface ocean carbonate system for seasonal to decadal studies of ocean acidification, Earth Syst. Sci. Data, 13, 777–808, https://doi.org/10.5194/essd-13-777-2021, 2021.

Gruber, N., Bakker, D. C. E., DeVries, T., Gregor, L., Hauck, J., Landschützer, P., McKinley, G. A., and Müller, J. D.: Trends and variability in the ocean carbon sink, Nat. Rev. Earth Environ., https://doi.org/10.1038/s43017-022-00381-x, 2023.

Halsey, L. G.: The reign of the p-value is over: what alternative analyses could we employ to fill the power vacuum?, Biol. Lett., 15, 20190174, https://doi.org/10.1098/rsbl.2019.0174, 2019.

Hauck, J., Nissen, C., Landschützer, P., Rödenbeck, C., Bushinsky, S., and Olsen, A.: Sparse observations induce large biases in estimates of the global ocean CO2 sink: an ocean model subsampling experiment, Philos. Trans. R. Soc. A Math. Phys. Eng. Sci., 381, 20220063, https://doi.org/10.1098/rsta.2022.0063, 2023.

Held, L. and Ott, M.: On p-values and Bayes factors, Annu. Rev. Stat. Appl, 5, 393–419, https://doi.org/10.1146/annurev-statistics-031017-100307, 2018.

Joos, F., Plattner, G.-K., Stocker, T. F., Marchal, O., and Schmittner, A.: Global Warming and Marine Carbon Cycle Feedbacks on Future Atmospheric CO2, Science, 284, 464–467, https://doi.org/10.1126/science.284.5413.464, 1999.

McKinley, G. A., Fay, A. R., Eddebbar, Y. A., Gloege, L., and Lovenduski, N. S.: External Forcing Explains Recent Decadal Variability of the Ocean Carbon Sink, AGU Adv., 1, e2019AV000149, https://doi.org/https://doi.org/10.1029/2019AV000149, 2020.

McNeil, B. I. and Matear, R. J.: The non-steady state oceanic CO2 signal: its importance, magnitude and a novel way to detect it, 10, 2219–2228, https://doi.org/10.5194/bg-10-2219-2013, 2013.

Lovenduski, N. S., Chatterjee, A., Swart, N. C., Fyfe, J. C., Keeling, R. F., and Schimel, D.: On the detection of COVID-driven changes in atmospheric carbon dioxide, Geophysical Research Letters, 48, e2021GL095396, https://doi.org/10.1029/2021GL095396, 2021

Orr, J. C., Kwiatkowski, L., and Pörtner, H.-O.: Arctic Ocean annual high in pCO2 could shift from winter to summer, Nature, 610, 94–100, https://doi.org/10.1038/s41586-022-05205-y, 2022.

Terhaar, J., Torres, O., Bourgeois, T., and Kwiatkowski, L.: Arctic Ocean acidification over the 21st century co-driven by anthropogenic carbon increases and freshening in the CMIP6 model ensemble, Biogeosciences, 18, 2221–2240, https://doi.org/10.5194/bg-18-2221-2021, 2021a.

Terhaar, J., Frölicher, T., and Joos, F.: Southern Ocean anthropogenic carbon sink constrained by sea surface salinity, Sci. Adv., 7, 5964–5992, https://doi.org/10.1126/sciadv.abd5964, 2021b.

Terhaar, J., Frölicher, T. L., and Joos, F.: Observation-constrained estimates of the global ocean carbon sink from Earth System Models, Biogeosciences, 19, 4431–4457, https://doi.org/10.5194/bg-19-4431-2022, 2022b.

Terhaar, J., Goris, N., Müller, J. D., DeVries, T., Gruber, N., Hauck, J., Perez, F. F., and Séférian, R.: Assessment of global ocean biogeochemical models for ocean carbon sink estimates in RECCAP2 and recommendations for future studies, Global Biogeochem. Cycles, 16, e2023MS003840, https://doi.org/10.1029/2023MS003840, 2024.

Yasunaka, S., Manizza, M., Terhaar, J., Olsen, A., Yamaguchi, R., Landschützer, P., Watanabe, E., Carroll, D., Adiwara, H., Müller, J. D., and Hauck, J.: An assessment of CO2 uptake in the Arctic Ocean from 1985 to 2018, Global Biogeochem. Cycles, 2023.

Zeng, J., Iida, Y., Matsunaga, T., and Shirai, T.: Surface ocean CO2 concentration and air-sea flux estimate by machine learning with modelled variable trends, Front. Mar. Sci., 9, 989233, https://doi.org/10.3389/fmars.2022.989233, 2022.