Response to RC2

We would like to express our gratitude to the reviewer for constructive and helpful comments/feedback. We have carefully addressed each question/comment and made changes where we agree that this would improve the manuscript. We have provided an itemized list below detailing our responses (in italic font) to the reviewer's suggestions.

Despite appreciating the author's efforts in this study, Reviewer has not been convinced by its originality. Based on ESM output, numerous existing research works have shown additional data sampling (e.g., bgcArgo, SOCCOM, Sailboat,...) critical for error reduction in pCO2 and flux estimation over the Southern Ocean and/or the global ocean [Bushinsky et al., 2019, Denvil-Sommer et al., 2021, Hauck et al., 2023, Landschützer et al., 2023]. One suggestion that would add value to the manuscript's findings is an analysis of spatial and temporal variations of flux estimates: to what extent their variability changes subject to the additional data. Some other major concerns are listed below.

Our study presents new findings that provide more insight into the number of additional samples and spatial pattern, consistent with current technology, that could reduce uncertainty in the ocean carbon sink, particularly in the Southern Ocean. There is no other study quantifying the impacts of meridional sampling by comparing different USV sampling tracks (also taking winter vs. summer sampling into account) in the Southern Ocean by using a Large Ensemble Testbed. Bushinsky et al. (2019) base their experiments on real-world SOCCOM float observations and use the SOM-FFN product for reconstruction. This is an important contribution. However, float-based estimates of pCO_2 are not incorporated into the SOCAT database and there are concerns about bias. It is therefore important to test the impact of realistic USV sampling, that can take direct pCO_2 observations with low uncertainties, can cover meridional gradients in the Southern Ocean, and are already incorporated into the SOCAT database.

The study by Hauck et al. (2023) uses GOBM output from one single model and reconstructs using two reconstruction methods (SOM-FFN and CarboScope), while we use ESM output from 75 different members and the pCO_2-Residual method. We also test a very different sampling pattern

compared to the "idealized" sampling in Hauck et al. (2023). We do find the study by Hauck et al. (2023) interesting, but note that it was not published when we submitted our initial manuscript. In the revised version we have added a paragraph discussing this study and comparing their results to ours. A key point made is that both Bushinsky et al. (2019) and Hauck et al. (2023) show an overestimation of the ocean sink with current sampling, while we show the opposite – an underestimation of the ocean sink. This further suggests that our study complements previous studies and adds value to this pertinent topic of ocean carbon research. It is important to present studies with different types of testbeds and reconstruction methods, so that we can better understand the impact of adding autonomous observations.

The study by Denvil-Sommer et al. (2021) is different to ours as it assesses sampling in the Atlantic Ocean, whereas our study focuses on sampling in the Southern Ocean and we show global reconstructions. Further, their study uses a different reconstruction method and assumes sampling from floats, not USVs.

Lines 149-153: "To build reconstruction algorithms through the data-driven training that occurs in ML, the statistics in all other algorithms developed to date must identify a function that disentangles these competing effects of SST on pCO2. Here, the algorithm is assisted by removing this known temperature effect, and it must therefore only learn the pCO2 impacts from biogeochemical drivers": there exist many other ML approaches [Friedlingstein et al., 2022] which do not separate the SSTeffects from others on pCO2 but succeeds in estimate pCO2. The major concerns are how to assess the uncertainty derived from SST effect removal and impacts on the experiment outputs.

Our study is not an evaluation of different ML approaches, but rather an assessment of how sampling impacts pCO_2 reconstructions. An evaluation of the method itself has already been performed by Bennington et al. (2022). They demonstrated that the pCO_2-Residual method performs better compared to other products when evaluating against independent data. They also showed improved skill when using pCO_2-Residual as the target variable as opposed to pCO_2. We want to assess how different sampling patterns affect the pCO_2 reconstruction. As we use the same

method for all experiments, we can directly compare them and evaluate how sampling impacts the reconstructions.

2. Figure 3: Relatively small bias and RMSE values have shown their imprints on the SOCAT track compared to "unseen" model truth. This evidences the problems of model overfitting. The authors can double-check whether model overfitting comes from the cross-validation technique or the pCO2-Residual method. As the key findings of this manuscript are based on the data reconstruction results, Reviewer suggests the authors to carefully verify their methods and solve the problems of model overfitting before further consideration for publication.

*We would argue that the global mean bias and RMSE for the SOCAT reconstruction is comparable to values shown for pCO_2 reconstructions using other methods (e.g., Stamell et al., 2020; Gregor et al., 2019). For example, as shown in **Figure 3**, bias generally ranges between -10 to +10 µatm, which is comparable to the study by Hauck et al. (2023). However, after carefully evaluating our calculations following the reviewer's feedback, we noticed an error in our code that calculates the RMSEs. After fixing this error, the mean RMSE values increased by ~ 3-4 µatm.*

Editorial and specific comments:

1. Lines 11-12: "anthropogenic" can be removed. The SO has taken up atmospheric CO2 without specifying natural or anthropogenic sources.

The Southern Ocean actively cycles natural and absorbs anthropogenic carbon. Gruber et al. (2009) demonstrate that the Southern Ocean is a source for natural carbon. The ocean sink for anthropogenic carbon is what we wish to focus on in this discussion.

2. Line 37: "fCO2" is not defined. "uncertainty of < 5 µatm": this holds only for the measurements chosen to provide gridded SOCAT datasets.

Noted and revised: "The Surface Ocean CO_2 ATlas (SOCAT; Bakker et al., 2016) is the largest global database of surface ocean CO_2 observations, with data starting in 1957. The main synthesis

and gridded products contain over 33 million high-quality direct shipboard measurements of fCO_2 (fugacity of CO_2) with an uncertainty of < 5 µatm (Bakker et al., 2022)".

3. Line 42: "Observation-based data products" −→ "Data mapping methods".

We wish to use the term 'observation-based data products' consistently following recent literature (e.g., Fay et al., 2021; Crisp et al., 2022; Friedlingstein et al., 2023).

4. Line 45: "These data products" −→ "These methods".

See above comment.

5. Lines 46-47: please remove or change ";" in the brackets to facilitate reading. You can use "-" instead. Line 47: "xCO2; atmospheric CO2" −→ "atmospheric CO2 - xCO2"

Noted and revised.

6. Line 48: "where these are co-located" −→ "where their available data are colocated".

We chose to keep the original sentence.

7. Lines 50-51: "Since the data products rely on observations to train the algorithms and thus produce these relationships": please rephrase this sentence. Data products do not train algorithms and produce relationships, but the ML-based methods themselves estimate the function between predictors and target data!

Noted and revised: "Since the data products rely on pCO_2 observations to estimate functions between the target and driver variables, data sparsity remains a fundamental limitation to this technique".

8. Line 57: "indirect pCO2 estimates": can you define this term? Are they computed from float measurements of other carbonate variables?

Noted and revised. We added this sentence: "These large uncertainties and biases arise when pCO_2 is not measured directly as in the observations included in SOCAT, but is rather estimated using measurements of pH combined with a regression-derived alkalinity estimate (Williams et al., 2017; Gray et al., 2018). SOCAT includes only direct pCO_2 observations".

9. Lines 67-68: "Such improvements in sampling are critically important in the undersampled Southern Ocean": USVs with low measurement uncertainty would prompt to be employed for observing network systems of pCO2 but to draw this statement, it requires to provide the availability of USVs to sample pCO2 by showing the sampling frequency and data coverage area over the SO?

Additional high-accuracy observations from the sparsely sampled Southern Ocean, such that can be obtained by USVs, are key to provide further constraints on the ocean carbon sink and air-sea flux. We do not believe it is necessary to go into detail about the data coverage over the Southern Ocean, as we reference studies such as Bakker et al. (2016, 2022) describing the SOCAT coverage (which includes the Saildrone observations from Sutton et al. (2021) in the latest version). We also mention that the SOCAT coverage is shown in supplementary ***Fig. S3*** *(****Fig. S5*** *in the revised version).*

10. Line 86: "actual observations": should be clarified. If you used the SOCAT grided data tracks in your LET experiments, please change to "SOCAT observation-based data" or "SOCAT gridded data".

We have revised the sentence: "However, instead of using real-world observations, we sample the target (i.e., surface ocean pCO_2) and driver variables (i.e., SST, SSS, MLD, Chl-a and xCO_2) from our Large Ensemble Testbed (LET) of Earth System Models (ESMs) (e.g., Stamell et al., 2020; Gloege et al., 2021; Bennington et al., 2022a)".

11. Lines 89-90: "in an ESM, surface ocean pCO2 is known at all times and locations": not precise enough. It depends on which approximations and computational resources. So far, the models have been derived at 1 ◦ or 0.25◦ and monthly resolutions?

We are just aiming to convey that an ESM will not have huge gaps like in the real ocean. We have revised the sentence: "First, in an ESM, the surface ocean pCO_2 field is provided precisely at all model times and 1°x1° points". The models used in our study have a 1°x1° resolution, which is stated multiple times throughout the manuscript.

12. Lines 161-162: "where pCO2 mean and SST mean is the long-term mean of surface ocean pCO2 and temperature, respectively, using all 1°x1° grid cells from the testbed": pCO2 mean is different regionally, why you don't compute a global map of pCO2 mean?

We do compute a mean of pCO_2 globally, which is the pCO_2^{mean} and this is used to calculate the residual.

13. Lines 165-168: Please clarify. The authors have excluded pCO2-Residual which have values below −250 µatm or over 250 µatm. They mention that such outliers correspond to model values higher than the maximum SOCAT data (816 µatm) and that do not reflect reality. It is not correct. First, both negative and positive pCO2- Residual values cannot represent the upper bound of SOCAT data. Second, SOCAT only covers a tiny portion of the global ocean at a monthly time scale, and there might exist unobserved pCO2 values higher than 816 µatm (e.g., over permanently or seasonally strong upwelling regions: Eastern Equatorial Pacific, Western Arabian Sea, Benguela, etc).

We are not saying that both negative and positive pCO_2-Residual values represent the upper bound of SOCAT data. Our statement is "These pCO_2-Residual values ***generally*** *correspond to high pCO_2, above the maximum value in SOCAT (816 µatm)". By this we mean that the majority of the pCO_2-Residual values that have been filtered out represent pCO_2 values that are larger than 816 µatm. However, since this seemed to be unclear, we have re-phrased this sentence: "Prior to algorithm processing, pCO_2-Residual values > 250 µatm and < -250 µatm from the testbed were*

filtered out targeting values that are not representative of the real ocean. The majority of the pCO_2-Residual values that were filtered out correspond to high pCO_2, above the maximum value in SOCAT (816 µatm; Stamell et al., 2020)".

14. Lines 310-311: "Our presentation of global maps is limited to runs 'x5_5Y_W' (5022 observations) and 311 'Z_x4_10Y_YR' (7600 observations)". The information of gridded data used in the experiments should be declared in addition to the number of observations by USVs.

We revised the sentence: "Our presentation of global maps is limited to runs 'x5_5Y_W' (5,022 monthly 1°x1° observations) and 'Z_x4_10Y_YR' (7,600 monthly 1°x1° observations)".

15. Lines 319-321: How did the authors compute Bias (and RMSE) over the global ocean? In order to fairly compare the results of two or more runs (e.g., zigzag vs one-latitude, SOCAT vs SOCAT+USV), error statistics are computed on modelbased data excluding all used in ML training. Specifically, the evaluation should not consider 'zigzag+one-latitude' ('SOCAT+USV') pCO2 data.

The reviewer is correct - the training data should ideally be removed before computing error statistics. When using actual observations, one would evaluate the reconstruction based on the test set alone. However, since we are using a model testbed, we have the opportunity to evaluate against pCO_2 values from "unseen" grid cells as well. In our study, we compute error statistics based on the full reconstruction, however this should have been communicated more clearly. The training data represents only about 1% of the full reconstruction. Below, we show the 75-member testbed spread in bias and RMSE calculated based on the full reconstruction (what we present in our study) vs. 'unseen' grid cells for the 'SOCAT-baseline'. The difference in mean bias and RMSE between the full and 'unseen' reconstruction is only 0.01 µatm and 0.08 µatm, respectively. The results from the different runs can therefore be compared even though the full reconstruction is taken into account. We agree however with the reviewer that the training data should have been removed. Considering that we would have to re-run all experiments, and it would not change the error statistics significantly or change our conclusions, we chose not to move forward with this for

this study. However, for future studies using the testbed, the training set will be removed before calculating statistical metrics.

*We now add mention of this: "Here, we calculate error statistics based on the full reconstruction (pCO_2 from all 1°x1° grid cells of the testbed, except for those masked or filtered out). In the full reconstruction, ~ 99 % of the data do not correspond to SOCAT or Saildrone USV observations used to train the algorithm (**Fig. S1**). Training data would ideally be removed before performance evaluation, but since the training data represent only ~ 1 %, the impact of not removing them is negligible (**Fig. S2**)". (**Figs. S1** and **S2** are shown below).*

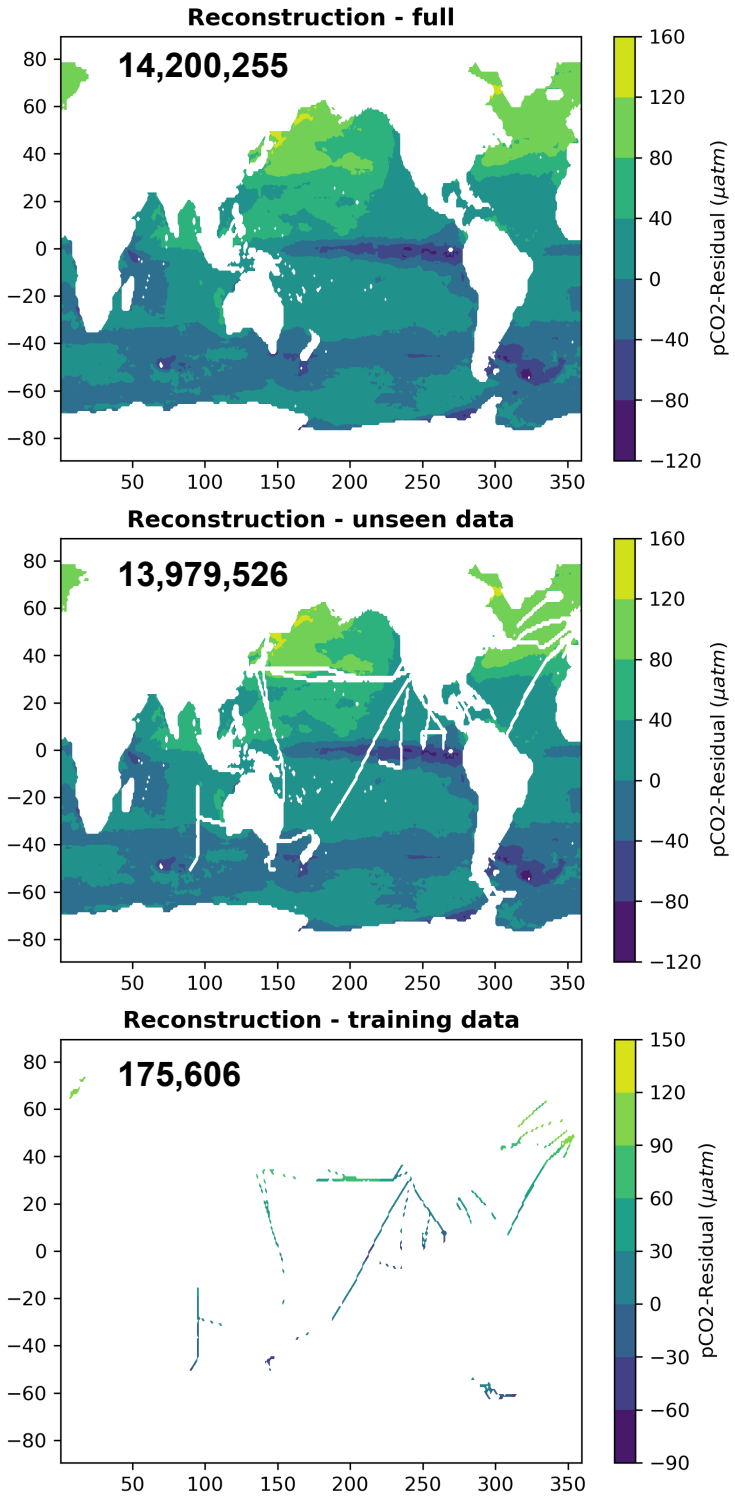


Figure S1*: Maps of the full pCO_2-Residual reconstruction (all 1°x1° grid cells of the testbed, except for those masked or filtered out; see* ***Section 2.1*** *and* ***2.2****), 'unseen' reconstruction (all 1°x1° grid cells that do not correspond to SOCAT observations), and training data from the testbed. The maps show data from CESM member 001 for the month of March 2016 for the 'SOCAT-baseline'.*

Numbers on panels represent the total monthly 1°x1° grid cells for the entire testbed period (1982-2016) for each group of data.

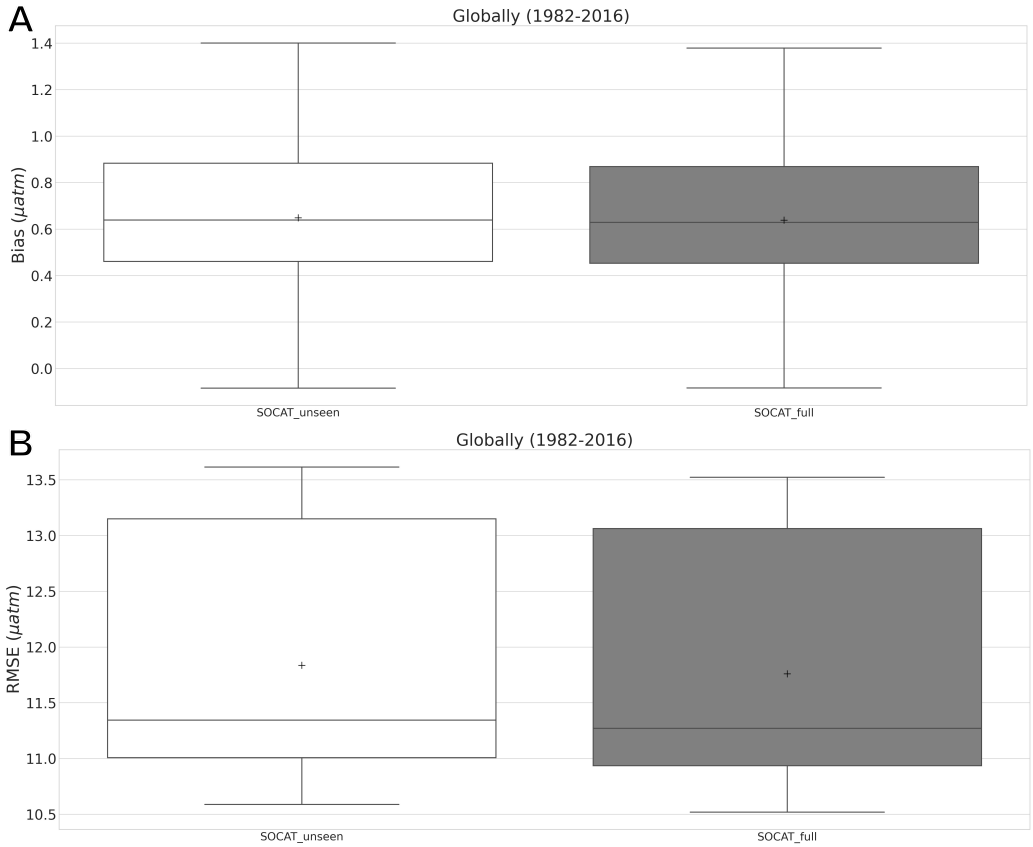
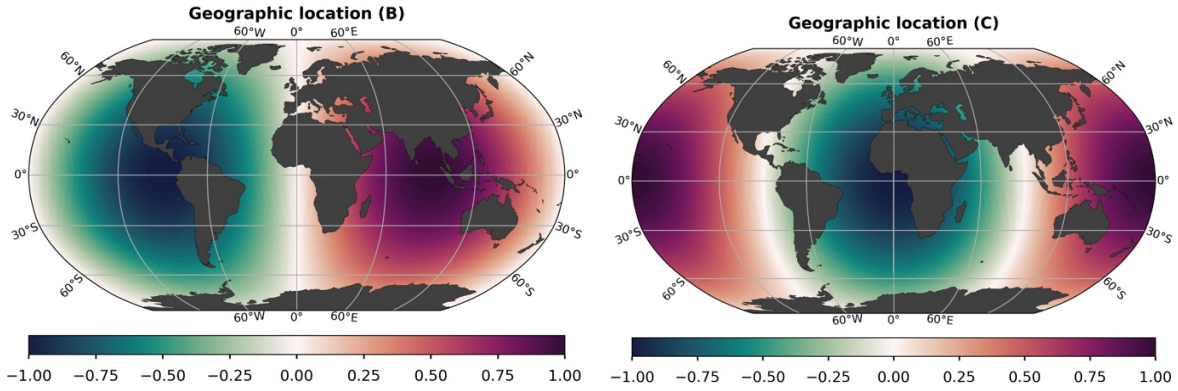


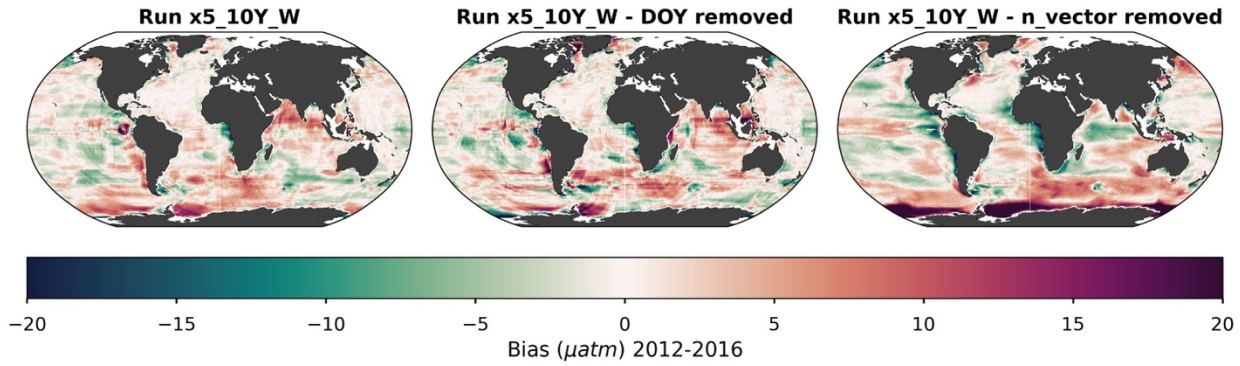
Figure S2*: Spread of bias (**a**) and RMSE (**b**) for the 75 members of the Large Ensemble Testbed for the 'unseen' and full reconstruction for the 'SOCAT-baseline'. The 'unseen' reconstruction represents independent data, i.e., all 1°x1° grid cells that do not correspond to SOCAT or Saildrone USV observations, and is not part of the training set.*

16. Figures S4 and S5 show cyclic marks (it would be exposed clearly if the authors use a discrete colormap with a low number of colors). Would they be imprints of a driver variable?

These "cyclic marks" are likely imprints of the three-component n-vector that replaces the longitude and latitude coordinates to continuous values between 0 and 1 (i.e., to avoid the algorithm interpreting 0 and 360 degrees to be far apart; see figure below).



Bennington et al. (2022) present global maps (their Fig. 4) of the feature importance of various driver variables used in the surface ocean pCO_2 reconstruction (MLD, SST, Chl-a, location and day of year). Such "cyclic marks" are apparent for "geographic location" and "day of year", but none of the other drivers. We did two test runs (using only one member from the testbed), removing day of year (DOY) and geographic location (n-vector; A, B and C) as inputs for the reconstruction. As shown by the figure below, the "cyclic" marks disappear when the n-vector is removed. When removing the n-vector transformation, however, the reconstruction shows significantly higher bias in the Southern Ocean, so we chose to keep these driver variables.



17. Figures 5 and 8: The author should report the number of data gridded from USV observations used in ML training. And the error statistics must be computed on the evaluation data (i.e., model-truth-based data excluding all the training data). Figure 8's caption: The mean of RMSEs here is computed with respect to space or time? Instead, the author should compute the mean of squared errors over the global ocean and the periods of interest and then report its square root.

*The number of monthly 1°x1° observations for each experiment is described in **Table 1** as well as shown on the x-axis of **Figure 5** and **8**. This was specified in the **Table 1** caption, but we now specify this in the figure captions as well: "'# additional observations' = number of monthly 1°x1° USV observations in addition to SOCAT". We state in the manuscript that: "The test and validation set each account for 20 % of the data, leaving 60 % for training". For both **Fig. 5** and **8**, the mean is computed with respect to both space (top figure shows global and bottom figure shows Southern Ocean, which in our study is defined as south of 35° S) and time, which is 2006-2016 (for the 10-year sampling) and 2012-2016 (for the five-year sampling). This is stated in the figure headlines.*

Regarding comment about error statistics, please see answer #15.

18. Line 386: ''Z_x10_5Y_YR

Noted and revised.

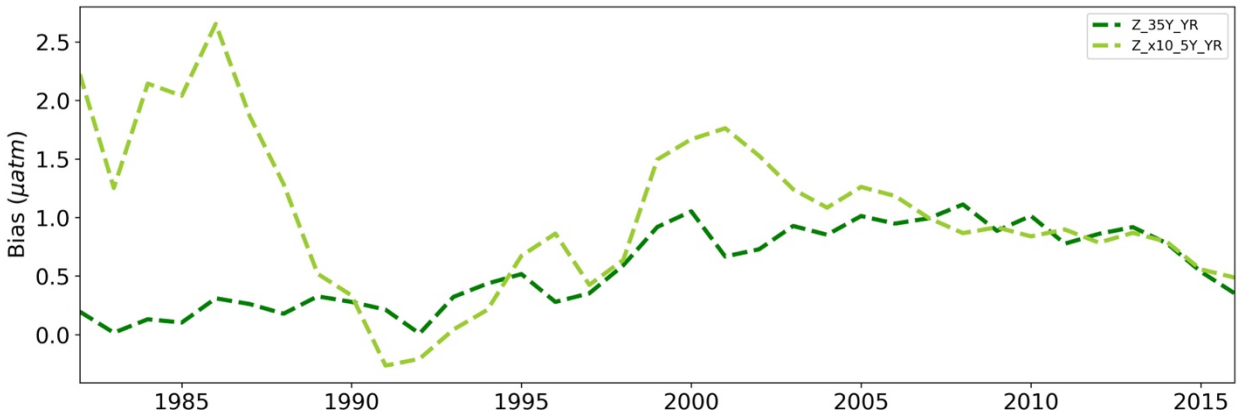
19. Lines 497-499: "Although run 'x13_10Y_W' demonstrates the highest reduction in bias out of all runs, the 'zigzag' runs still reduce bias in the Southern Ocean by 44-65 % (vs. 77 % for run 'x13_10Y_W')". The evaluation should not put high confidence on the bias reduction since this statistic is computed as the mean of negative and positive differences between pCO2 estimates and model truth. Reviewer agrees that the bias can be used to assess model over- or underestimation but RMSD is a better metric for an overall evaluation.

We agree with the reviewer, and that is why we report both bias and RMSE. Our conclusions do not fully rely on bias alone, as is shown throughout the paper. For example, we conclude that the zigzag-runs perform best overall, even though run 'x13_10Y_W' demonstrates a higher reduction with respect to mean bias.

20. Lines 536-541: "To better understand this discrepancy, we performed an additional experiment based on run 538 'Z_x10_5Y_YR', but assumed sampling every year for the entire testbed period (i.e., 1982-2016). The results from this experiment show a significant eduction in the temporal

variability of reconstruction bias; with the additional USV sampling, the reconstructed Southern Ocean air-sea CO2 flux closely matches the 'model truth' for the entire testbed duration (Fig. S14).". Here biases increases in the last two decades that do not reflect the increase in the number of SOCAT (SOCAT+USV) data as shown in the previous results.

*As shown by the figure below, run 'Z_x10_**5Y**_YR' (shown in **Fig. 6** in main text) and 'Z_x10_**35Y**_YR' (shown in **Fig. S14** in supplement; in the revised version, this is now **Fig. S20**) show similar variability the last five years when the sampling is identical. For run Z_x10_**5Y**_YR', USV observations have been added only for the last five years of the testbed, while for run 'Z_x10_**35Y**_YR', USV observations have been added for the whole testbed period (35 years). The bias decreases more significantly in the earlier decades for run 'Z_x10_**5Y**_YR' because there are no additional USV observations at this time, and there are significantly less SOCAT observations in this period compared from 1990 and onwards (see **Fig. S3c**; in the revised version, this is now **Fig. S5c**).*



21. Lines 552-554: "Further, we find that this modest amount of additional Saildrone USV sampling increases the global and Southern Ocean air-sea CO2 flux by up to 0.1 Pg C yr-1, 25% of the uncertainty in the ocean carbon sink ". The increase in global ocean CO2 sink estimated by the LET testbed can not be compared with the uncertainty derived from the GCB's quantification [Friedlingstein et al., 2022]. First, they are two different statistics. Second, the GCB's uncertainty is computed based on the ensemble of different data mapping and modeling methods, and thus the value might be significantly larger than the one estimated by each method itself.

These values can be compared as they are in the same units. We wish to demonstrate that 0.1 Pg C/yr is a significant reduction. Following the reviewer's comment, we revised the sentence: "Further, we find that this modest amount of additional Saildrone USV sampling increases the global and Southern Ocean air-sea CO_2 flux by up to 0.1 Pg C yr^{-1}, a quantity equivalent to 25 % of the uncertainty in the ocean carbon sink".

References

Bakker, D. C. E., Pfeil, B., Landa, C. S., Metzl, N., O'Brien, K. M., Olsen, A., Smith, K., Cosca, C., Harasawa, S., Jones, S. D., Nakaoka, S., Nojiri, Y., Schuster, U., Steinhoff, T., Sweeney, C., Takahashi, T., Tilbrook, B., Wada, C., Wanninkhof, R., Alin, S. R., Balestrini, C. F., Barbero, L., Bates, N. R., Bianchi, A. A., Bonou, F., Boutin, J., Bozec, Y., Burger, E. F., Cai, W.-J., Castle, R. D., Chen, L., Chierici, M., Currie, K., Evans, W., Featherstone, C., Feely, R. A., Fransson, A., Goyet, C., Greenwood, N., Gregor, L., Hankin, S., Hardman-Mountford, N. J., Harlay, J., Hauck, J., Hoppema, M., Humphreys, M. P., Hunt, C. W., Huss, B., Ibánhez, J. S. P., Johannessen, T., Keeling, R., Kitidis, V., Körtzinger, A., Kozyr, A., Krasakopoulou, E., Kuwata, A., Landschützer, P., Lauvset, S. K., Lefèvre, N., Lo Monaco, C., Manke, A., Mathis, J. T., Merlivat, L., Millero, F. J., Monteiro, P. M. S., Munro, D. R., Murata, A., Newberger, T., Omar, A. M., Ono, T., Paterson, K., Pearce, D., Pierrot, D., Robbins, L. L., Saito, S., Salisbury, J., Schlitzer, R., Schneider, B., Schweitzer, R., Sieger, R., Skjelvan, I., Sullivan, K. F., Sutherland, S. C., Sutton, A. J., Tadokoro, K., Telszewski, M., Tuma, M., van Heuven, S. M. A. C., Vandemark, D., Ward, B., Watson, A. J., and Xu, S.: A multi-decade record of high-quality fCO_2 data in version 3 of the Surface Ocean CO_2 Atlas (SOCAT), Earth System Science Data, 8, 383–413, https://doi.org/10.5194/essd-8-383-2016, 2016.

Bakker, D. C. E., Alin, S. R., Becker, M., Bittig, H. C., Castaño-Primo, R., Feely, R, A., Gkritzalis, T., Kadono, K., Kozyr, A., Lauvset, S, K., Metzl, N., Munro, D, R., Nakaoka, S., Nojiri, Y., O'Brien, K, M., Olsen, A., Pfeil, Benjamin, P., Denis, S., Tobias, S., Kevin F., Sutton, A. J., Sweeney, C., Tilbrook, B., Wada, C., Wanninkhof, R., Willstrand W. A., Akl, J., Apelthun, L. B., Bates, N., Beatty, C. M., Burger, E. F., Cai, W., Cosca, C. E., Corredor, J. E., Cronin, M., Cross, J. N., De Carlo, E. H., DeGrandpre, M. D., Emerson, S. R., Enright, M. P., Enyo, K., Evans, W., Frangoulis, C., Fransson, A., García-Ibáñez, M. I., Gehrung, M., Giannoudi, L., Glockzin, M., Hales, B., Howden, S. D., Hunt, C. W., Ibánhez, J. S. P., Jones, S. D., Kamb, L., Körtzinger, A., Landa, C. S., Landschützer, P., Lefèvre, N., Lo Monaco, C., Macovei, V. A., Maenner J. S., Meinig, C., Millero, F. J., Monacci, N. M., Mordy, C., Morell, J. M., Murata, A., Musielewicz, S., Neill, ., Newberger, T., Nomura, D., Ohman, M., Ono, T., Passmore, A., Petersen, W., Petihakis, G., Perivoliotis, L., Plueddemann, A. J., Rehder, G., Reynaud, T., Rodriguez, C., Ross, A. C., Rutgersson, A., Sabine, C. L., Salisbury, J. E., Schlitzer, R., Send, U., Skjelvan, I., Stamataki, N., Sutherland, S. C., Sweeney, C., Tadokoro, K., Tanhua, T., Telszewski, M., Trull, T., Vandemark, D., van Ooijen, E., Voynova, Y. G., Wang, H., Weller, R. A., Whitehead, C., Wilson, D.: Surface Ocean CO_2 Atlas Database Version 2022 (SOCATv2022) (NCEI Accession 0253659), NOAA National Centers for Environmental Information [dataset], https://doi.org/10.25921/1h9f-nb73, 2022.

Bennington, V., Galjanic, T., and McKinley, G. A.: Explicit Physical Knowledge in Machine Learning for Ocean Carbon Flux Reconstruction: The pCO_2-Residual Method, Journal of Advances in Modeling Earth Systems, 14(10), https://doi.org/10.1029/2021ms002960, 2022.

Bushinsky, S. M., Landschützer, P., Rödenbeck, C., Gray, A. R., Baker, D., Mazloff, M. R., Resplandy, L., Johnson, K. S., and Sarmiento, J. L.: Reassessing Southern Ocean air-sea CO_2 flux estimates with the addition of biogeochemical float observations, Global Biogeochemical Cycles, 33(11), 1370-1388, https://doi.org/10.1029/2019GB006176, 2019.

Crisp, D., Dolman, H., Tanhua, T., McKinley, G. A., Hauck, J., Bastos, A., Sitch, S., Eggleston, S., & Aich, V.: How well do we understand the land-ocean-atmosphere carbon cycle? Reviews of Geophysics, 60(2), e2021RG000736, doi:10.1029/2021RG000736, 2022.

Denvil-Sommer, A., Gehlen, M., & Vrac, M.: Observation system simulation experiments in the Atlantic Ocean for enhanced surface ocean pCO_2 reconstructions, Ocean Science, 17(4), 1011-1030, https://doi.org/10.5194/os-17-1011-2021, 2021.

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 CO_2 fluxes from surface pCO_2 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., Jarníková, 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., Smallman, T. L., Smith, S. M., Sospedra-Alfonso, R., Sun, Q., Sutton, A. J., Sweeney, C., Takao, S., Tans, P. P., Tian, H., Tilbrook, B., Tsujino, H., Tubiello, F., van der Werf, G. R., van Ooijen, E., Wanninkhof, R., Watanabe, M., Wimart-Rousseau, C., Yang, D., Yang, X., Yuan, W., Yue, X., Zaehle, S., Zeng, J., and Zheng, B.: Global Carbon Budget 2023, Earth Syst. Sci. Data, 15, 5301–5369, https://doi.org/10.5194/essd-15-5301-2023, 2023.

Gloege, L., McKinley, G. A., Landschützer, P., Fay, A. R., Frolicher, T. L., and Fyfe, J. C.: Quantifying Errors in Observationally Based Estimates of Ocean Carbon Sink Variability, Global Biogeochemical Cycles, 35(4), https://doi.org/10.1029/2020gb006788, 2021.

Gray, A. R., Johnson, K. S., Bushinsky, S. M., Riser, S. C., Russell, J. L., Talley, L. D., Wanninkhof, R., Williams, N. L., and Sarmiento, J. L.: Autonomous biogeochemical floats detect significant carbon dioxide outgassing in the high-latitude Southern Ocean, Geophysical Research Letters, 45(17), 9049-9057, https://doi.org/10.1029/2018GL078013, 2018.

Gregor, L., Lebehot, A. D., Kok, S., and Monteiro, P. M. S.: A comparative assessment of the uncertainties of global surface ocean CO_2 estimates using a machine-learning ensemble (CSIR-ML6 version 2019a) – have we hit the wall?, Geoscientific Model Development, 12, 5113-5136, https://doi.org/10.5194/gmd-12-5113-2019, 2019.

Gruber, N., gloor, M., Mikaloff Fletcher, S. E., Doney, S. C., Dutkiewicz, S., Folows, M. J., Gerber, M., Jacobson, A. R., Joos, F., Lindsay, K., Menemenlis, D., Mouchet, A., Muller S. A., Sarmiento, J. L., & Takahashi, T.: Oceanic sources, sinks, and transport of atmospheric CO_2, Global Biogeochemical Cycles, 23(1), https://doi.org/10.1029/2008GB003349, 2009.

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: and ocean model subsampling experiment, Philosophical Transactions Of the Royal Society A, 381:20220063, https://doi.org/10.1098/rsta.2022.0063, 2023.

Stamell, J., Rustagi, R. R., Gloege, L., and McKinley, G. A.: Strengths and weaknesses of three Machine Learning methods for pCO_2 interpolation, Geoscientific Model Development Discussions[preprint], doi:10.5194/gmd-2020-311, 22 October 2020.

Sutton, A. J., Williams, N. L., and Tilbrook, B.: Constraining Southern Ocean CO_2 flux uncertainty using uncrewed surface vehicle observations, Geophysical Research Letters, 48(3), e2020GL091748, https://doi.org/10.1029/2020GL091748, 2021.

Takahashi, T., Sutherland, S.C., Wanninkhof, R., Sweeney, C., Feely, R.A., Chipman, D.W., Hales, B., Friederich, G., Chavez, F., Sabine, C. and Watson, A.: Climatological mean and decadal change in surface ocean pCO2, and net sea–air CO2 flux over the global oceans. Deep Sea Research Part II: Topical Studies in Oceanography, 56(8-10), pp.554-577, 2009.

Williams, N. L., Juranek, L. W., Feely, R. A., Johnson, K. S., Sarmiento, J. L., Talley, L. D., Dickson, A. G., Gray, A. R., Wannikhof, R., Russell, J. L., Riser, S. C., and Takeshita, Y.: Calculating surface ocean pCO_2 from biogeochemical Argo floats equipped with pH: An uncertainty analysis, Global Biogeochemical Cycles, 31(3), 591-604, https://doi.org/10.1002/2016GB005541, 2017.