Supplementary for "Explicit physical knowledge in machine learning for ocean carbon flux reconstruction: The pCO_2-Residual Method"

Val Bennington[1,2], Tomislav Galjanic[3], Galen A McKinley[1]

[1]Lamont-Doherty Earth Institute, Columbia University

[2]Makai Ocean Engineering, Hawaii

[2]Data Science Institute, Columbia University

Contents of this file

1. Text S1 to S5

2. Figures S1 to S9

S1. Large Ensemble Testbed Findings

Utilizing the Large Ensemble Testbed (Gloege et al., 2021), we analyzed how RMSE was impacted by reconstructing the pCO_2-Residual using the new technique instead of utilizing the algorithm with pCO_2 as the target variable (Section 2.1.2). The algorithm with pCO_2 as the target is that which is used to calculate the long-term mean pCO_2 for input to the calculation of pCO_2-T (Section 2.1). The Large Ensemble Testbed consists of 25 ensemble members each from 4 Earth System Models, 100 members total. Within the Testbed, we sample model features and pCO_2 at the same times and locations as we have actual SOCAT observations, in every ensemble member. Just as done with actual observations, an XGBoost algorithm is trained on the subset of features and pCO_2 from the models. We then reconstruct pCO_2 everywhere using the resulting functions and compare the reconstructed pCO_2 to the model "truth". Thus, the reconstructed pCO_2 can be evaluated at all times and locations, not just where we have SOCAT observations. Figure S1 shows that in addition to reducing the RMSE of the test data for each reconstruction ("test data"), RMSE across the globe, where the model has never been sampled ("unseen data"), is reduced using the pCO_2-Residual approach. Note also that against both test and unseen data, the high extreme RMSE is reduced by at least 3 μatm.

Since we know the model truth everywhere, at all times within the Testbed, we can evaluate bias, mean absolute error (MAE), and how well variability on different timescales is captured by both techniques. Figure S2 shows the mean bias and mean absolute error (MAE) over the 38 year period within the Testbed. The pCO_2-Residual technique reduces and bias and mean absolute error within the subtropical oceans. The ability to capture sub-decadal and decadal variability is substantially improved over the use of pCO_2 alone as the target variable (Figure S3, S4). Comparison to LDEO and GLODAP data over separate decades (Figure S5, (Gloege et al., 2022)) also indicates good performance of pCO_2-Residual on decadal timescales.

S2. Uncertainty Due to Chlorophyll Climatology

Within the Large Ensemble Testbed, we use XGBoost to reconstruct pCO_2 using time-varying chlorophyll-a (every month has modeled chlorophyll-a). We compare to when the monthly climatology of model chlorophyll (1998 onward) is used for prior to 1998. As we do not have satellite observations of chlorophyll-a prior to 1997, this technique is used to estimate uncertainties caused by using a climatology of chlorophyll-a for the years prior to satellite observations. The calculated air-sea CO_2 flux differs significantly prior to the mid-1990s and decreases to approximately 0.05 PgC/yr by 2005 (Figure S6). There is variation across the models, with the largest mean impact on the reconstruction seen within the MPI model. The mean difference across the ESMs and time is less than 0.1 PgC/yr.

S3. RMSE, Bias, MAE in pCO_2-Residual approach

The map of mean RMSE against all SOCAT observations using the pCO_2-Residual algorithm is shown in Figure S7. We see lowest RMSE in temperature-controlled subtropical regions, with values less than 10 μatm, as expected, and higher RMSE outside of these regions.

S4. CO_2 flux across ensemble members of pCO_2-Residual

Figure S8 and S9 demonstrate that there are minimal differences in CO_2 flux across the ensemble members of pCO_2-Residual (Table 3).

S5. Test of Clustering with Self-Organizing Maps

To examine whether the regression would be improved by dividing the global ocean into biomes, we utilized the self-organizing map package SOMPY (Moosavi et al., 2014) (https://github. com/sevamoo/SOMPY). The global ocean was divided into 5, 10, and 15 clusters using maximum annual ice fraction, mean pCO_2, mean annual sea surface temperature, mixed layer depth, and spring mean chlorophyll (Fay & McKinley, 2014). On the global scale, there was no added skill, quantified based on RMSE and comparisons to independent data at BATS, HOT, LDEO, or GLODAP. We therefore maintain the simpler model.

References

Fay, A. R., & McKinley, G. A. (2014). Global open-ocean biomes: mean and temporal variability. *Earth System Science Data*, *6*, 273–284. doi: doi:10.5194/essd-6-273-2014

Gloege, L., McKinley, G. A., Landschützer, P., Fay, A. R., Frölicher, T. L., Fyfe, J. C., ... Takano, Y. (2021). Quantifying errors in observationally based estimates of ocean carbon sink variability. *Global Biogeochemical Cycles*, *35*(4), e2020GB006788. Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2020GB006788` (e2020GB006788 2020GB006788) doi: https://doi.org/10.1029/2020GB006788

Gloege, L., Yan, M., Zheng, T., & McKinley, G. A. (2022). Improved quantification of ocean carbon uptake by using machine learning to merge global models and pCO_2 data. *Journal of Advances in Modeling Earth Systems*, *14*. doi: 10.1029/2021MS002620

Moosavi, V., Packmann, S., & Vallés, I. (2014). *Sompy: A python library for self organizing map (som).* (GitHub.[Online]. Available: https://github.com/sevamoo/SOMPY)

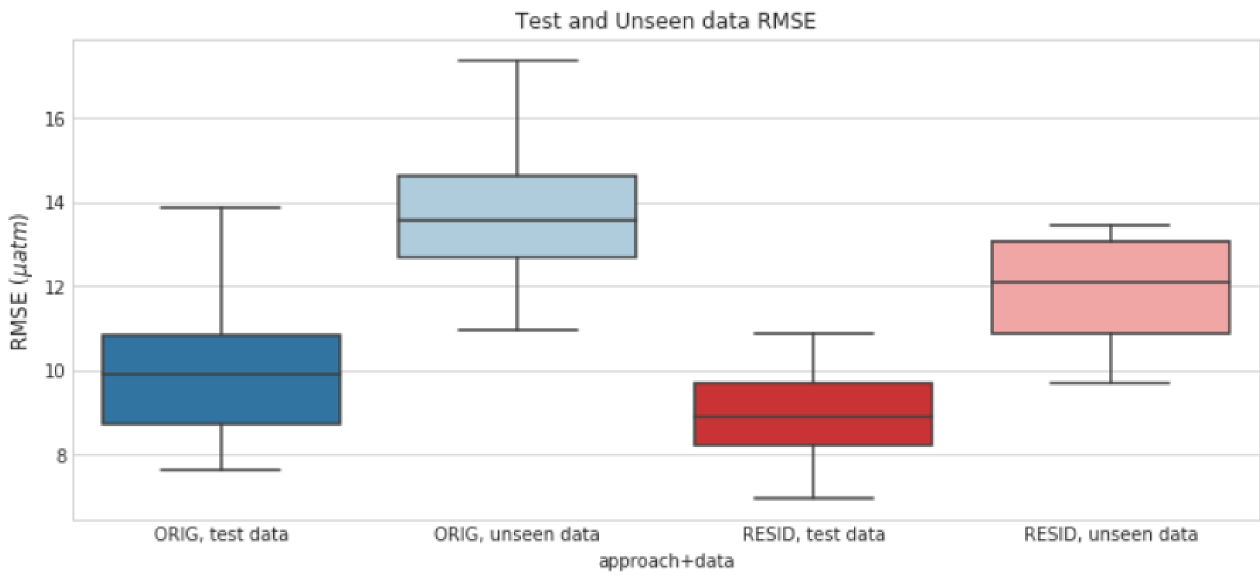

Figure S1. Test RMSE for direct pCO_2 reconstruction (ORIG, test data); RMSE at locations not sampled by SOCAT for direct pCO_2 reconstruction (ORIG, unseen data); Test RMSE for pCO_2-Residual approach (RESID, test data); RMSE at locations not sampled for pCO_2-Residual approach (RESID, unseen data). Each boxplot contains the 100 ensemble members, 25 from each Earth System Model of the Large Ensemble Testbed (Gloege et al., 2021).

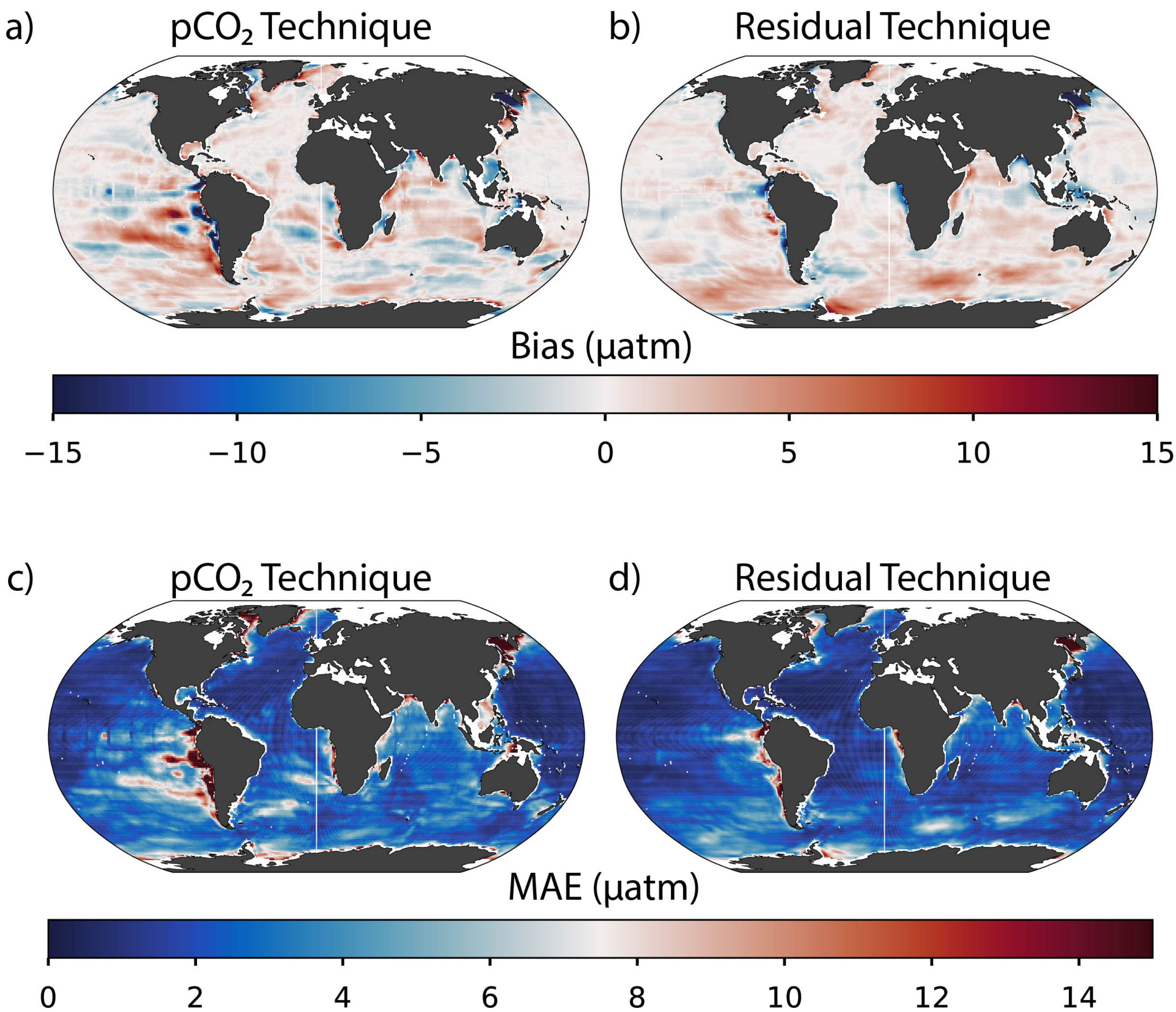

Figure S2. Mean bias of the typical approach reconstructing pCO_2 (a) and using the Residual technique (b) within the Large Ensemble Testbed (Gloege et al., 2021). Mean absolute error within the Testbed when reconstructing pCO_2 directly (c) and reconstructing pCO_2 using the Residual technique (d).

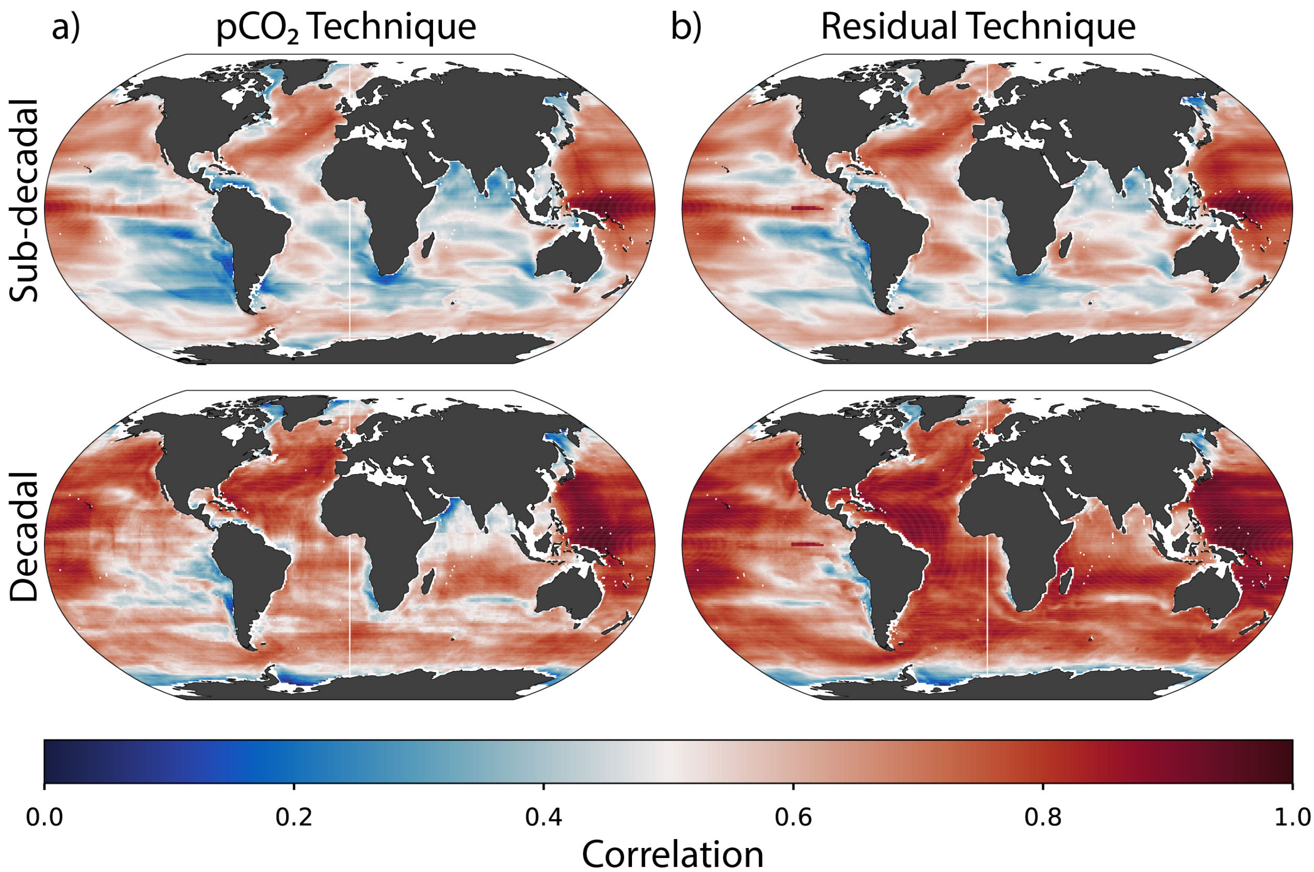

Figure S3. Correlations between reconstructed pCO_2 and the Testbed truth pCO_2 on sub-decadal (top row) and decadal timescales (bottom row). Correlations for the approach where pCO_2 is the target variable on the left. Correlations for the pCO_2-Residual approach on the right.

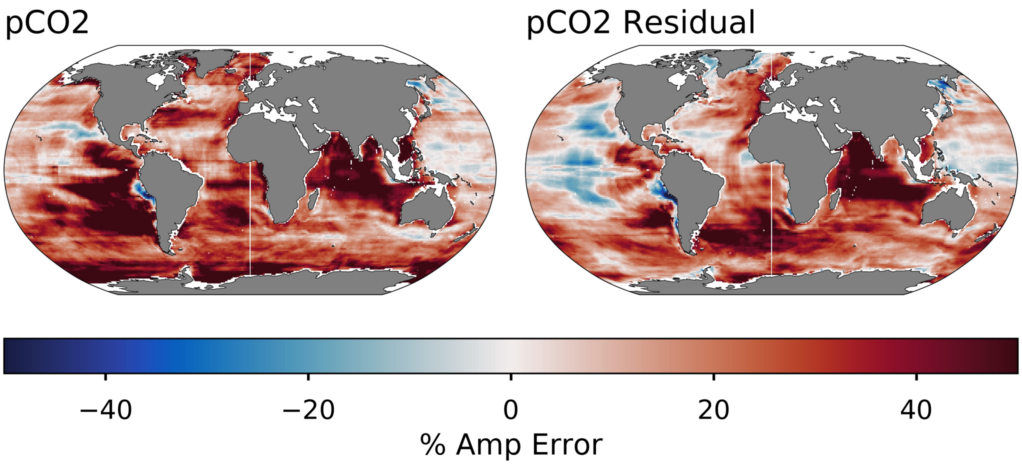

Figure S4. Percent error in the decadal variability amplitude within the Large Ensemble Testbed (Gloege et al., 2021): with pCO_2 directly reconstructed (Section 2.1.2) on left and with new pCO_2-Residual approach on right. Zero percent error is optimal.

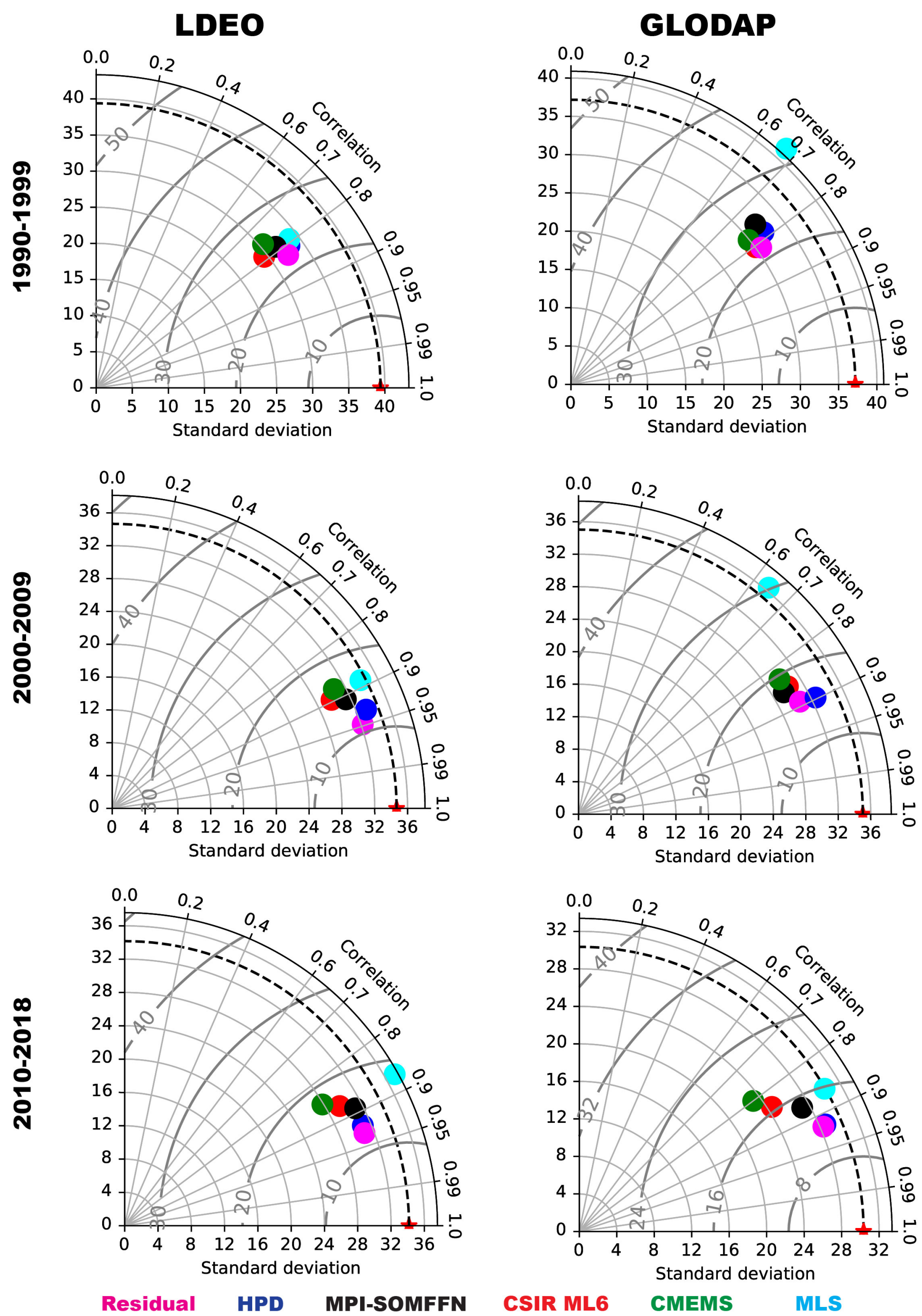

Figure S5. Following Gloege et al. (2022), Taylor diagrams of correlation, standard deviation, and bias-adjusted RMSE for the decades 1990-1999 (top), 2000-2009 (middle), and 2010-2018 (bottom) compared to pCO_2 from LDEO (left) and GLODAP (right). See section 2.5 text for details on these independent datasets.

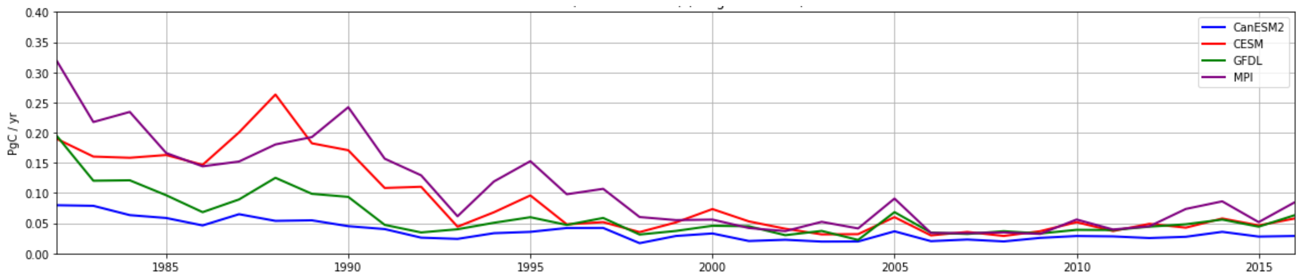

Figure S6. Absolute value of difference in globally-integrated reconstructed CO_2 flux (PgC/yr) when using a climatology of chlorophyll-a prior to 1998. Different colors represent the four different ESMs of the Large Ensemble Testbed (Gloege et al., 2021).

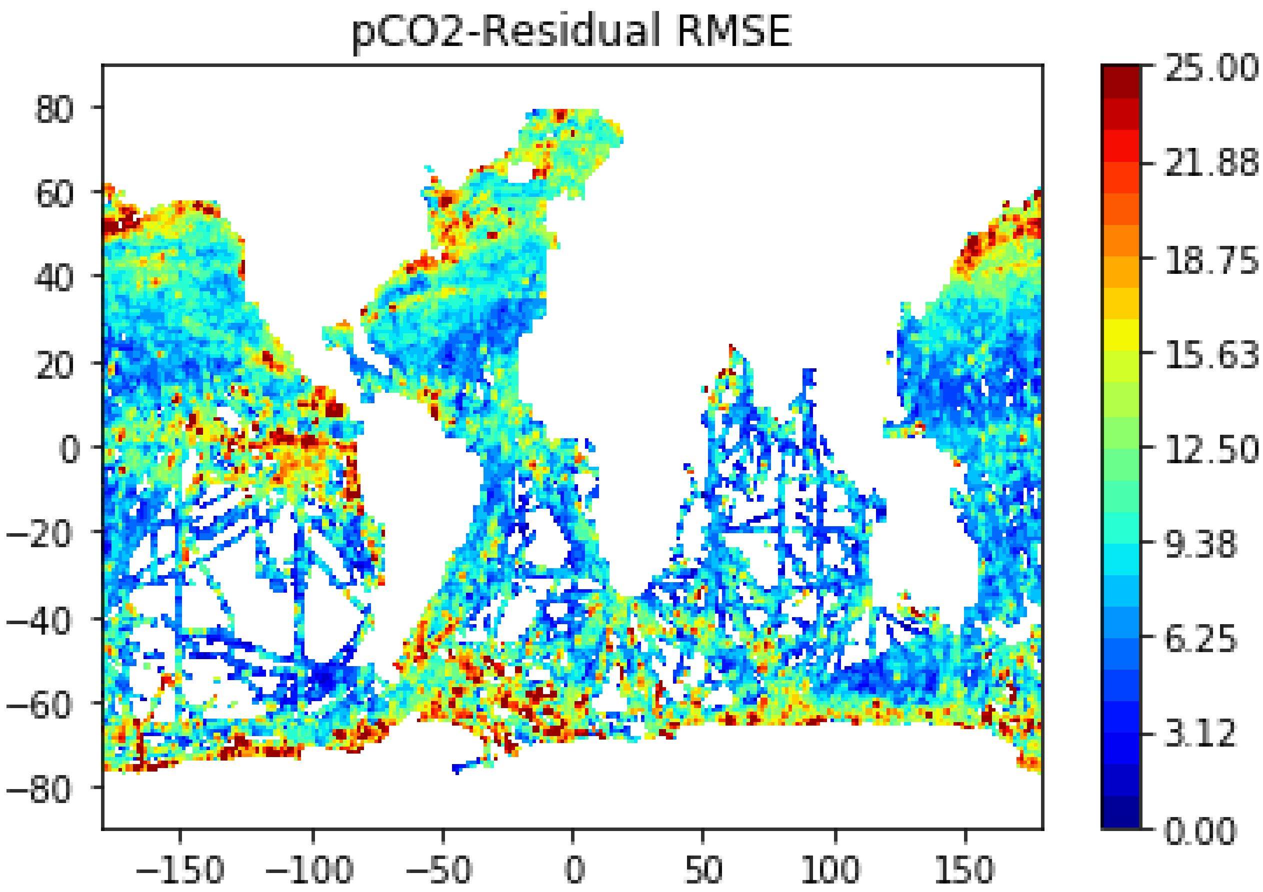

Figure S7. Mean RMSE (μatm) across the global ocean using the pCO_2-Residual approach.

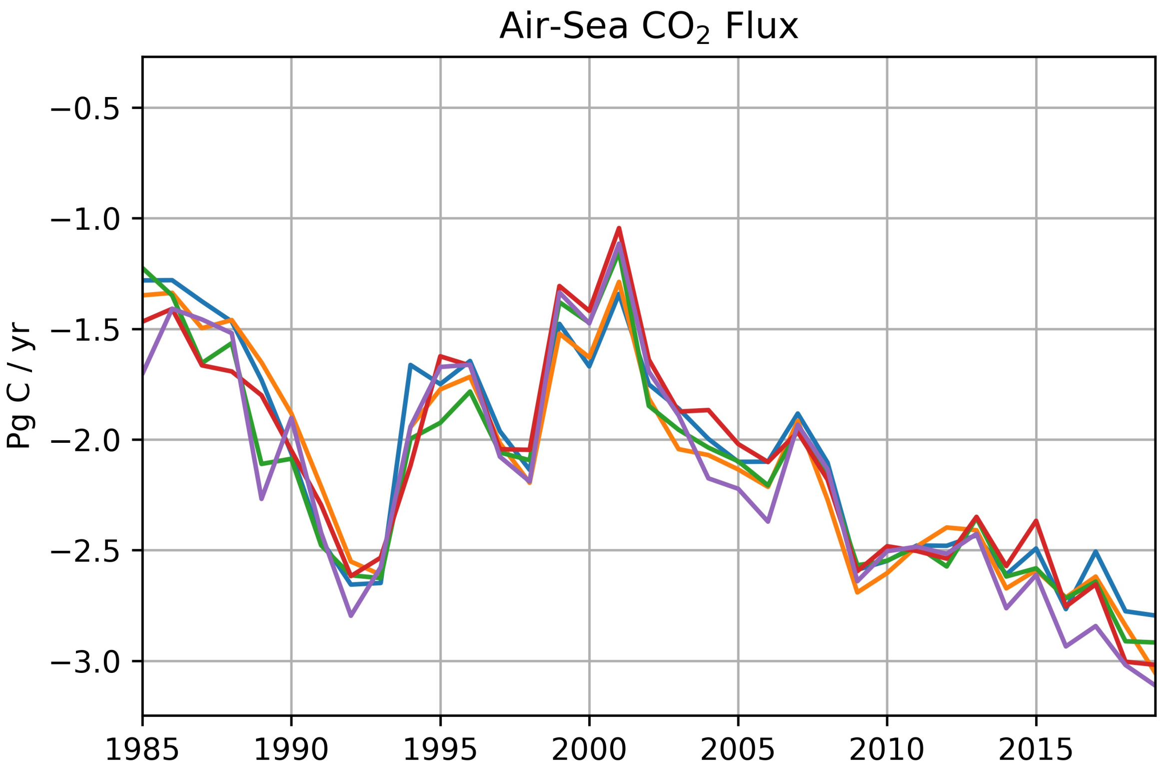

Figure S8. Air-sea CO_2 fluxes (PgC/yr) for each of the individual ensemble members of the pCO_2-Residual reconstruction (Table 3).

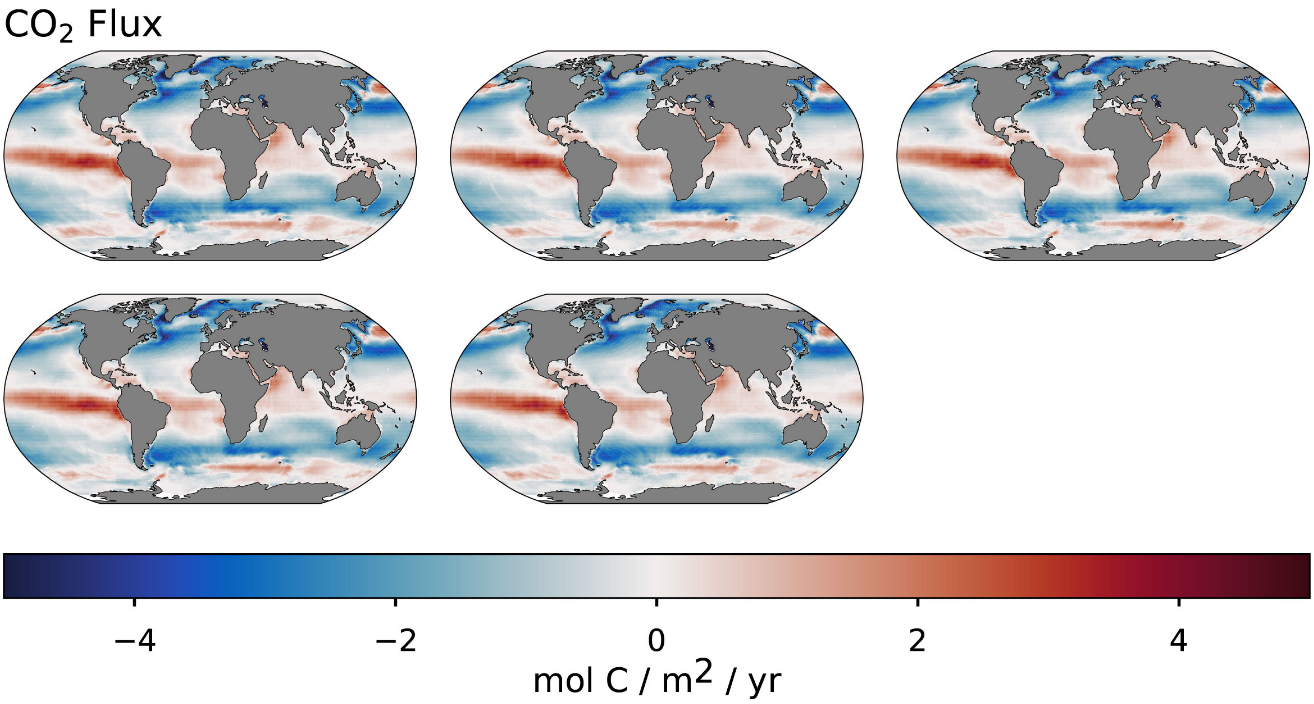

Figure S9. Mean air-sea CO_2 flux for 1985-2019 (molC/m^2/yr) for each of the five ensemble members of the pCO_2-Residual reconstruction (Table 3). There are minimal differences in the estimates.