# Improved Quantification of Ocean Carbon Uptake by Using Machine Learning to Merge Global Models and pCO$_2$ Data

**L. Gloege[1,2]** , **M. Yan[3]**, **T. Zheng[3,4]** , and **G. A. McKinley[1]**

[1]Lamont-Doherty Earth Observatory and Department of Earth and Environmental Sciences, Columbia University, New York, NY, USA, [2]Now at NASA-GISS, New York, NY, USA, [3]Department of Statistics, Columbia University, New York, NY, USA, [4]Data Science Institute, Columbia University, New York, NY, USA

**Abstract** The ocean plays a critical role in modulating climate change by sequestering $CO_2$ from the atmosphere. Quantifying the $CO_2$ flux across the air-sea interface requires time-dependent maps of surface ocean partial pressure of $CO_2$ ($pCO_2$), which can be estimated using global ocean biogeochemical models (GOBMs) and observational-based data products. GOBMs are internally consistent, mechanistic representations of the ocean circulation and carbon cycle, and have long been the standard for making spatio-temporally resolved estimates of air-sea $CO_2$ fluxes. However, there are concerns about the fidelity of GOBM flux estimates. Observation-based products have the strength of being data-based, but the underlying data are sparse and require significant extrapolation to create global full-coverage flux estimates. The Lamont Doherty Earth Observatory-Hybrid Physics Data (LDEO-HPD) $pCO_2$ product is a new approach to estimating the temporal evolution of surface ocean $pCO_2$ and air-sea $CO_2$ exchange. LDEO-HPD uses machine learning to merge high-quality observations with state-of-the-art GOBMs. We train an eXtreme Gradient Boosting (XGB) algorithm to learn a non-linear relationship between model-data mismatch and observed predictors. GOBM fields are then corrected with the predicted model-data misfit to estimate real-world $pCO_2$ for 1982–2018. The resulting reconstruction by LDEO-HPD is in better agreement with independent $pCO_2$ observations than other currently available observation-based products. Within uncertainties, LDEO-HPD global ocean uptake of $CO_2$ agrees with other products and the Global Carbon Budget 2020.

**Plain Language Summary** The ocean absorbs carbon from the atmosphere, which slows climate change. In order to estimate how much carbon the ocean absorbs, we need to know how much is exchanged from the atmosphere into the ocean at each location over time. The direct observations required to do this are very sparse and in some regions of the ocean, observations have never been made. One approach to fill in the gaps is to use machine-learning techniques, which are algorithms that build a relationship for ocean carbon based on related satellite observations with global coverage. Another approach is to use computer simulations, which use mathematical equations to represent ocean processes. Here, we merge these two innovations by blending model output with machine-learning to create a hybrid product: the Lamont Doherty Earth Observatory-Hybrid Physics Data (LDEO-HPD). Particularly for the most recent decade, LDEO-HPD agrees better with independent observations than other products, indicating promise for the approach.

## 1. Introduction

The ocean's net uptake of $CO_2$ is a key component of the global carbon cycle. Quantifying how anthropogenic emissions are distributed between atmosphere, land biosphere, and ocean reservoirs with as low uncertainty as possible is needed to support international climate policy (Peters et al., 2017). The Global Carbon Budget 2020 (Friedlingstein et al., 2020) finds that for 2009–2018, the ocean sink for anthropogenic carbon was $-2.5 \pm 0.6$ PgC/yr (negative flux into the ocean), based on global ocean biogeochemical models (GOBMs). However, four observation-based products suggest a trend diverging from the GOBMs over this period and a sink that is 0.4 PgC/yr larger (Hauck et al., 2020). Do the relatively new observation-based estimates indicate an issue with the long-used GOBMs?

Estimating the global ocean $CO_2$ sink requires knowledge of ocean partial pressure of $CO_2$ ($pCO_2$). The Surface Ocean $CO_2$ ATlas (SOCAT) is an annually compiled database of surface ocean fugacity of $CO_2$ ($fCO_2$) with over 28.2 million observations for 1957–2019 in the SOCATv2020 release (Bakker et al., 2016), mainly from volunteer observing ships. $fCO_2$ is nearly equivalent to $pCO_2$, different by a 0.3% non-ideality correction; we make

**Resources:** L. Gloege, G. A. McKinley
**Software:** L. Gloege, M. Yan
**Supervision:** T. Zheng, G. A. McKinley
**Validation:** L. Gloege
**Visualization:** L. Gloege
**Writing – original draft:** L. Gloege, G. A. McKinley
**Writing – review & editing:** L. Gloege, M. Yan, T. Zheng, G. A. McKinley

this adjustment in our analysis to derive $pCO_2$ (Section 2.1). Due to limited number of ships, routes, and the high cost of maintenance, the data retrieved from this observation system remains sparse in space and time. Data are concentrated in the Northern Hemisphere (Figures 1a–b1c). Using these data alone, $pCO_2$ cannot be quantified at all times and all locations, and thus statistical extrapolations have been performed to create observation-based data products (Rödenbeck et al., 2015). The carbon cycle community uses these products along with GOBMs (Friedlingstein et al., 2020) to independently estimate $CO_2$ fluxes and, through their analysis and comparison, to improve knowledge of the global ocean carbon cycle. We propose an explicit merging of the strengths of both approaches in the form of a hybrid observation-based data product that uses $pCO_2$ estimates from multiple GOBMs as a prior.

A GOBM is a knowledge-based model that parameterizes the physical, chemical, and biological processes influencing surface ocean $pCO_2$ using a system of coupled differential equations. GOBMs have long been taken as the best estimate of the anthropogenic air-sea $CO_2$ flux, and have always been the basis for quantification of the ocean carbon sink in the annual Global Carbon Budget published by the Global Carbon Project since 2009 (Friedlingstein et al., 2020; Le Quéré et al., 2009). Nine GOBMs were used as the basis for the Global Carbon Budget in 2019 and 2020 (Friedlingstein et al., 2019, 2020). These models are certainly imperfect, with substantial differences among them and potentially an underestimation of $CO_2$ flux variability, particularly in the Southern Ocean (Gloege et al., 2021; Gruber, Landschützer, et al., 2019; Hauck et al., 2020). Regional mean flux biases can also be substantial (Fay & McKinley, 2021). However, based on a long history of their application to understanding and quantification of air-sea $CO_2$ fluxes, it is a sensible to use GOBMs as a prior estimate upon which data-based improvements can be made.

Most observation-based products find a relationship between a suite of datasets and the target variable (ocean $pCO_2$) using machine learning algorithms. The statistical relationships of the algorithm are dependent on the quantity and quality of SOCAT $pCO_2$ data, driver data, and the skill of the reconstruction algorithm. A recent assessment of the SOM-FFN (Landschützer et al., 2014) reconstruction indicates high fidelity for the mean and seasonality of $pCO_2$-based $CO_2$ flux estimates. However, $pCO_2$ data sparsity (Figures 1a–b1c) limits the ability to reconstruct interannual to decadal timescale variations (Gloege et al., 2021). Though the spread across the full suite of recently published products is smaller than the spread across the current generation of models (McKinley et al., 2020), there remain substantial differences in the timing and amplitude of interannual variability (Friedlingstein et al., 2020). In a comprehensive evaluation of multiple products, Gregor et al. (2019) find comparable skill with respect to independent data in the current generation of products, and suggest that we have reached a skill limit for these products that is fundamentally due to data sparsity.

Both GOBMs (Friedlingstein et al., 2020; Hauck et al., 2020) and observation-based products (Rödenbeck et al., 2015) provide approximately global estimates of ocean $pCO_2$ and $CO_2$ flux. The two approaches differ significantly in the way they estimate ocean $pCO_2$. GOBMs compute the evolution of physical and biogeochemical processes based on complex systems of coupled differential equations that can only be solved numerically. Observation-based products do not explicitly incorporate known physics, but instead estimate a non-linear relationship between a suite of driver variables (e.g., see Table 2) and ocean $pCO_2$ where these are co-located. Global full-coverage driver datasets are then processed through these relationships to estimate global full-coverage $pCO_2$. GOBMs and observation-based data products generally agree on the large-scale patterns and long-term increase in ocean $pCO_2$ (Landschützer et al., 2014; McKinley et al., 2016, 2020; Tjiputra et al., 2014). GOBMs have comparable root mean square errors against SOCAT $pCO_2$ to those in the observation-based products when compared over large spatial regions, indicating comparable skill (Gregor et al., 2019; Hauck et al., 2020). However, GOBMs are biased high when sub-sampled at SOCAT observation locations (Figure 1d). In some models, this global bias is at least partially attributable to the exclusion of the well-established water vapor correction (Dickson et al., 2007) in the calculation of atmospheric $pCO_2$ (McKinley et al., 2020).

As noted above, the $pCO_2$ data typically available to train machine learning algorithms are spatially sparse (Figures 1a–b1c). Data availability also changes over time (Figures 2a and 2b). This trend in data availability, combined with the long-term positive trend in ocean $pCO_2$ (~33 $\mu$atm increase from 1980s to 2010s) has the potential to impact the ability of algorithms to represent the data. Machine learning, or any statistical fit, performs best when target variables distributions have the same shape as the driver variables (Goodfellow et al., 2016). With ocean $pCO_2$ as the target variable, the algorithm is being asked to predict a broad and right-skewed distribution (Figure 2a) that is unlike the driver variables that do not have a significant trend across time. For other products,
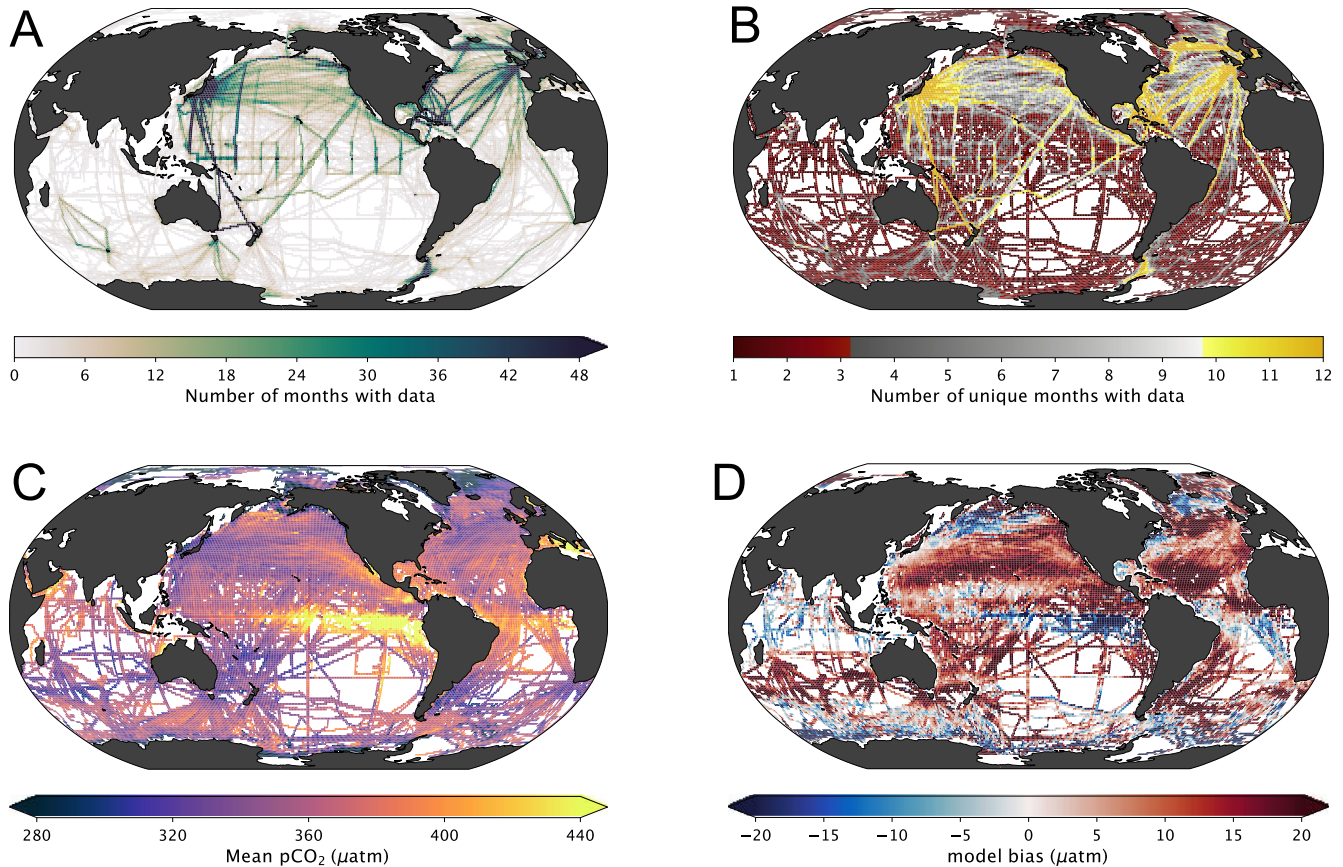
**Figure 1.** (a) Total number of months over 1982–2018 with observations. (b) Number of unique months with observations. (c) Long-term mean $pCO_2$ at each $1° \times 1°$ pixel. (d) Bias between Surface Ocean $CO_2$ ATlas and mean of nine global ocean biogeochemical models.

the temporal trend of $pCO_2$ is addressed by including atmospheric $pCO_2$ as a predictor variable. We note that the difference between observed and GOBM-estimated $pCO_2$ has only a modest long-term trend (~9 $\mu$atm from 1980s to 2010s, Figures 2c and 2d). Thus, if we use the difference between SOCAT observations and the GOBMs as a basis for algorithm development, we largely address the aforementioned concern. In other words, with model-data misfit as our target variable, the skewness of the target variable is substantially reduced (Figure 2c).

In this study, we leverage the nine GOBMs used in the Global Carbon Budget 2020 (Friedlingstein et al., 2020) and combine them with a supervised machine learning algorithm to create the Lamont Doherty Earth Observatory-Hybrid Physics Data ocean $pCO_2$ observation-based product (LDEO-HPD). Instead of using ocean $pCO_2$ as the target variable, as do other data products (Denvil-Sommer et al., 2019; Gregor et al., 2019; Landschützer et al., 2014; Rödenbeck et al., 2015), the target variable for our eXtreme Gradient Boosting (XGB) algorithm is the misfit between SOCAT observed $pCO_2$ and each model where SOCAT observations exist in space and time ($pCO_{2,SOCAT} - pCO_{2,GOBM}$). Our driver data are the same suite of in situ and satellite observations used by other approaches (see Table 2). To make final estimates of actual ocean $pCO_2$, the XGB algorithm first uses full-field observed driver data to predict model misfit at all locations for each GOBM. These misfit fields are then added back to each GOBM to make the final estimate. Each GOBM is processed using a unique algorithm, and the final LDEO-HPD output is the average of the nine merged data-model estimates. See Figure 3 for a schematic. Our approach of combining data-based machine learning with the physics embodied in dynamical models follows on recent innovations in physics-guided machine learning (Karpatne et al., 2017; Reichstein et al., 2019) and the use of machine learning to correct dynamical models (Watt-Meyer et al., 2021) for earth science applications.

A potential additional application of the approach we develop here is to use model-data misfit fields to visualize and quantify errors in GOBM carbon cycle simulations at broader temporal and spatial scales than is currently possible with actual SOCAT data (Hauck et al., 2020). Spatio-temporal misfit mapped by the algorithm is a direct

**Table 1**
*Reference for GOBMs Used in the Global Carbon Budget 2020 (Friedlingstein et al., 2020)*

| Global ocean biogeochemical models (GOBMs) | Reference |
|---|---|
| NEMO-PlankTOM5 | Buitenhuis et al. (2013) |
| MICOM-HAMOCC (NorESM1-OCv1.2) | Schwinger et al. (2016) |
| MPIOM-HAMOCC6 (MPI) | Paulsen et al. (2017) |
| NEMO3.6-PISCESv2-gas (CNRM) | Berthet et al. (2019) |
| CISRO | Law et al. (2017) |
| FESCOM-1.4-REcoM2 | Hauck et al. (2020) |
| MOM6-COBALT (princeton) | Adcroft et al. (2019) |
| CESCM-ETHZ | Doney et al. (2009) |
| NEMO-PISCES (IPSL) | Aumont et al. (2015) |

estimate of GOBM skill for locations where in situ data do not exist. We briefly explore this application in Section 3.1.

## 2. Methods

GOBM output is incorporated into a supervised machine learning algorithm to create a hybrid data product for 1982–2018. We use gradient boosting as implemented in the XGB library (Chen & Guestrin, 2016). XGB learns a non-linear relationship between a suite of features and the misfit between the GOBM and direct SOCAT observations. We use this approach to upscale SOCAT $pCO_2$ observations and create a nearly global, temporally complete data product. The upscaled $pCO_2$ product is statistically evaluated against independent observations and other published data products. A schematic of LDEO-HPD is shown in Figure 3. From $pCO_2$ estimated with LDEO-HPD, we estimate $CO_2$ flux using the standard bulk parameterization that relates the flux to wind speed (Fay et al., 2021; Wanninkhof, 1992, 2014).

### 2.1. Pre-Processing SOCAT Observations

We use surface ocean $pCO_2$ calculated from the SOCAT v2019 monthly gridded $fCO_2$ product. SOCAT v2019 is a quality-controlled data set that contains observations of surface ocean $fCO_2$, which is converted to $pCO_2$ with Equation 1,

$$pCO_2 = fCO_2 \cdot \exp\left( P_{atm}^{surf} \cdot \frac{B + 2\delta}{R \cdot T} \right)^{-1} \tag{1}$$

where $P_{atm}^{surf}$ is the atmospheric surface pressure from ECMWF Reanalysis version 5, $T$ is the sea surface temperature (SST) in Kelvin from National Oceanic and Atmospheric Administration (NOAA) optimally interpolated SST version 2 (OISSTv2), $B$ and $\delta$ are virial coefficients from Weiss (1974), $R$ is the gas constant (Dickson et al., 2007).

### 2.2. Global Ocean Biogeochemical Models

As a first guess for ocean $pCO_2$, we use output from nine GOBMs (Table 1) which participated in the Global Carbon Budget 2020 (Friedlingstein et al., 2020), with the final year being 2018. Meteorological reanalysis and atmospheric $CO_2$ are used to force each model (Hauck et al., 2020). Each GOBM parameterizes the physical, chemical, and biological processes influencing surface ocean $pCO_2$ using a system of coupled differential equations. The surface $pCO_2$ from each GOBM is bi-linearly interpolated from the native model grid to a $1° \times 1°$ monthly resolution to be consistent with SOCAT gridded observations (Sabine et al., 2013).

### 2.3. Machine Learning Method and the LDEO-HPD Product

XGB (Chen & Guestrin, 2016) is a supervised machine learning algorithm where multiple features, $X$, are used to predict a target variable $y$. The XGB algorithm can then be used to estimate a function, $f(X)$, such that: $y \approx f(X)$. The algorithm begins with an initial guess for $y$, a choice to which the algorithm is not sensitive. As illustrated in Figure 3b, a decision tree is used to learn the difference between the training data and the initial guess. This new tree is added to the initial guess. This process of adding trees to correct the errors made in the summation of previous trees is repeated until either a predefined number of trees has been made, or when adding an additional tree results in no further improvement. Improvements are marginal if there are more than about 1,000 predefined trees. A significantly lower number of predefined trees renders the method more sensitive. The final prediction is the sum of all trees such that the closest fit of input data and algorithm output is achieved. A mean-squared-error (MSE) loss function is minimized using gradient descent.

Gradient boosting algorithm, as implemented in the XGB library version 0.9 with the scikit-learn wrapper (Chen & Guestrin, 2016), is used to find a non-linear relationship between a suite of input features and the misfit

**Table 2**
*Summary of the Products, Variables, and Data Processing Steps Used for Feature and Target Variables*

| Group: product | Variable | Abbreviation | Processing |
|---|---|---|---|
| SOCATv2019[a] | Partial pressure of ocean $CO_2$ | $pCO_2$ | See Section 2.1 |
| NOAA:OISSTv2[b] | Sea Surface Temperature | SST | - |
| | SST seasonal anomaly | SST' | SST - annual average |
| | Sea Ice Fraction | ICE | - |
| Met Office:EN4[c] | Sea Surface Salinity | SSS | - |
| NOAA:GLOBALVIEW[d] | Atmospheric $CO_2$ mixing ratio | $xCO_2$ | - |
| DeBoyer:Mixed Layer Depth[e] | Mixed Layer Depth | MLD | $\log_{10}(MLD)$ |
| ESA:GlobColour[f] | Chlorophyll-a | Chl a | $\log_{10}(Chla)$ |
| | Chl a seasonal anomaly | Chl a' | chl a - annual average |
| - | Day of year | $J_1$ | $\sin\left(\frac{j*2\pi}{365}\right)$ |
| | | $J_2$ | $\cos\left(\frac{j*2\pi}{365}\right)$ |
| - | n-vector | A | $\sin(\lambda)$ |
| | | B | $\sin(\mu)\cos(\lambda)$ |
| | | C | $-\cos(\mu)\cos(\lambda)$ |

*Note.* Data processing is described in the text. Symbol next to each product identifies the source.
[a]Reference: Bakker et al. (2016) Source: https://www.socat.info/. [b]Reference: Reynolds et al. (2002) Source: https://www.esrl.noaa.gov/psd/data/gridded/data.noaa.oisst.v2.html. [c]Reference: Good et al. (2013) Source: https://www.metoffice.gov.uk/hadobs/en4/. [d]Reference: Conway et al. (1994) Source: https://gml.noaa.gov/ccgg/mbl/. [e]Reference: de Boyer Montégut et al. (2004) Source: http://www.ifremer.fr/cerweb/deboyer/mld/home.php. [f] Reference: Maritorena et al. (2010) Source: http://www.globcolour.info/.

between each GOBM and SOCAT $pCO_2$: $(pCO_{2,SOCAT} - pCO_{2,GOBM})$. This algorithm was chosen because it leads to a better fit to input data than the other options considered, neural network or random forest (Stamell et al., 2020). To estimate $pCO_2$ at each spatial location, the algorithm relies on datasets with full, or approximately full, global coverage (Table 2): SST and Surface Chlorophyll-a (Chl-a) from satellite; Sea Surface Salinity (SSS) from a compilation of in-situ data sources (See Good et al., 2013 for a discussion of the data sources); Mixed layer depth (MLD) climatology from Argo floats; and atmospheric $CO_2$ mixing ratio ($xCO_2$) from station sites. These variables serve as proxies for known processes affecting $pCO_2$. Solubility is set by SSS and SST. Biological uptake of dissolved inorganic carbon (DIC) is indicated by Chl-a. Biological productivity and entrainment of DIC are influenced by MLDs. The long-term growth of ocean $pCO_2$ is driven by atmospheric $xCO_2$. Additional annual mean anomaly features are derived for SST and Chl-a by subtracting the annual mean from each year. These features help the algorithm learn more complex relationships and capture intra-annual variability. N-vector transformation (Gade, 2010; Gregor et al., 2017; Sasse et al., 2013) of latitude and longitude is included to help the algorithm learn spatial relationships. Although latitude and longitude do not have a direct mechanistic link to $pCO_2$, they are found to improve the reconstruction. This is attributable to these variables being proxies for biogeographic properties that structure $pCO_2$ fields (Fay & McKinley, 2013, 2017). Time transformation of the day of year constrains seasonality.

The features and associated $pCO_2$ misfit are split into three sets: validation, training, and testing. The test set is chosen as seven randomly selected years (1984, 1987, 1992, 1999, 2002, 2005, and 2014), these years account for about 16% of the data; thus we reserve approximately 20% of data for testing, as is standard in machine learning applications. The withheld test set is used to evaluate performance on a completely independent data set, individual years are withheld for the test set to retain individual ship tracks and increase the independence of test data from training and validation data (Gregor et al., 2019). The remaining data is randomly split between the validation and training set. The validation set, which accounts for 20% of the remaining data, is used to optimize the algorithms hyperparameters. The hyperparameters define the architecture of decision trees used in the model. The remaining data defines the training set, which is used to construct the decision trees (Gregor et al., 2019).
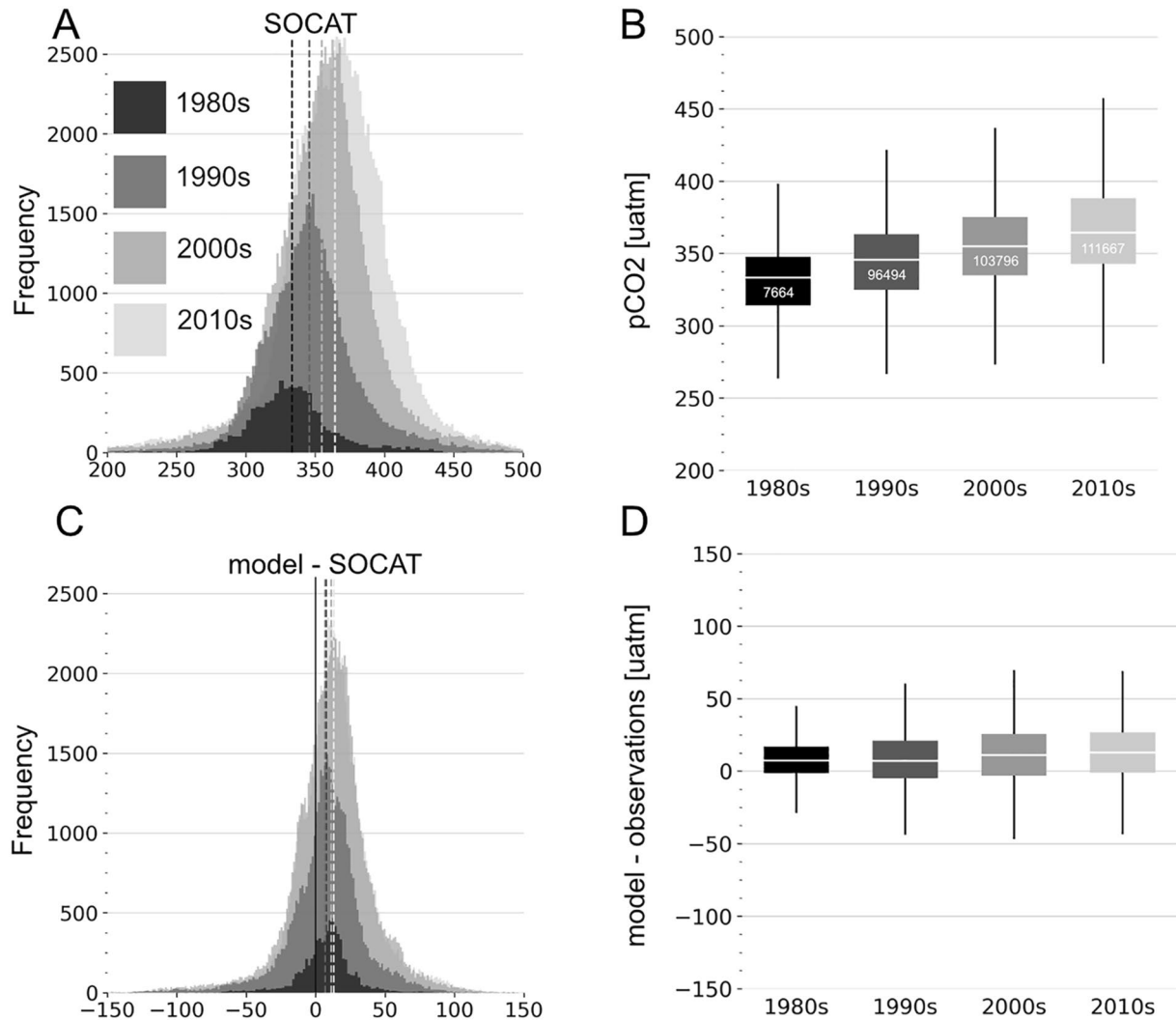
**Figure 2.** (a) Histogram of Surface Ocean CO$_2$ ATlas (SOCAT) pCO$_2$ observations in 1980s, 1990s, 2000s and 2010s shown by different shades of gray. Dotted line indicates mean pCO$_2$. (b) Boxplot of observations for each decade. Whisker indicates 1.5*IQR, observations outside the whisker have been omitted. White line indicates the mean and the number inside in the box indicates the number of observations within that decade. (c) Histogram of the difference between CESM model and SOCAT and (d) is the corresponding boxplot. Due to different internal model structures, the long-term trend from 1980s to 2010s varies from −7 $\mu$atm to +9 $\mu$atm.

Our XGB algorithm uses 1,500 decision trees each with a max depth of nine levels or until no further splits to the samples in that node are possible. Each new tree uses 95% of the features and a random subsample of 85% of the observations from the training set with replacement. The weight of each sequential tree is reduced by 5%. Light L1 regularization was applied to control overfitting and loss is measured using MSE.

XGB is used to estimate spatio-temporal estimates of the misfit for each of nine GOBMs. Misfit estimates at all locations in space and time are added back to the original GOBM to correct the GOBM toward the data. This process is repeated for each of the nine GOBMs. The final result is then the average of all nine predictions. A schematic of HPD is shown in Figure 3.

### 2.4. Independent Data Sets

Observations not included in the SOCAT database are used to validate the method (Table 3). These datasets include the Lamont-Doherty Earth Observatory (LDEO) database, with SOCAT data removed; and GLobal Ocean Data Analysis Project version 2 (GLODAPv2). Two time series sites are also used for validation: Bermuda Atlantic Time-series Study (BATS) and Hawaii Ocean Time-series (HOT). In these datasets, pCO$_2$ is either
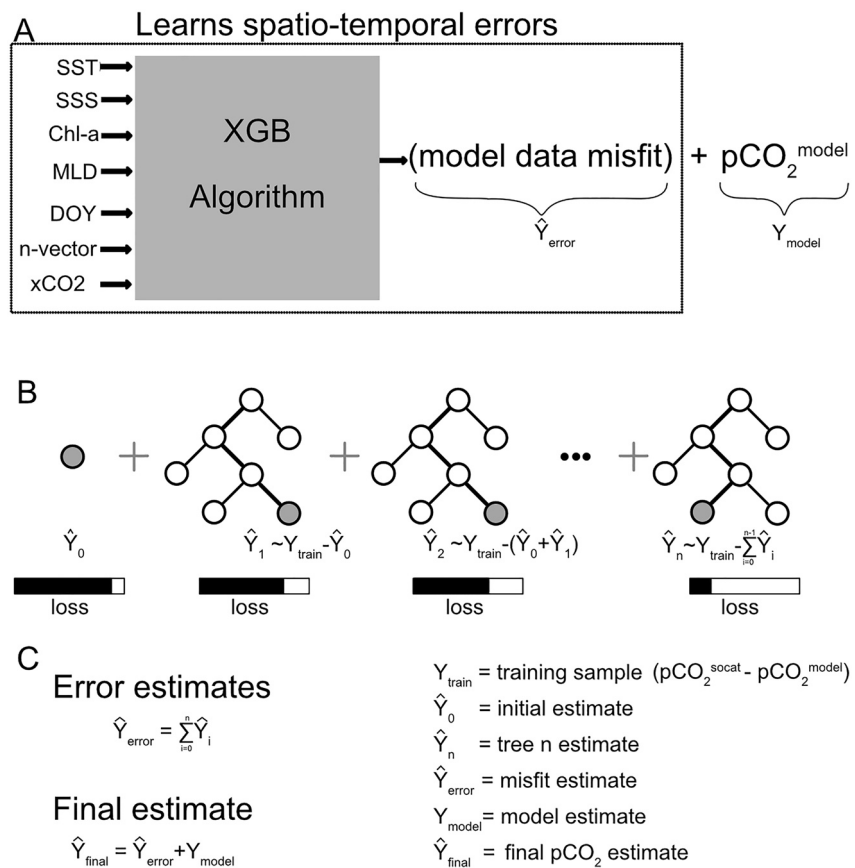
**Figure 3.** (a) Schematic of Lamont Doherty Earth Observatory-Hybrid Physics Data method. A relationship between a suite of auxiliary features and the model data misfit is learned via the eXtreme Gradient Boosting (XGB) algorithm. Spatio-temporal errors are then added back to the model's $pCO_2$ field to create the final product. (b) Outlines the XGB algorithm, where decision trees are sequentially added to improve the mistakes of the previous trees. Each additional tree reduces the loss and improves the overall performance of the algorithm. (c) The final estimate of $pCO_2$ is the model-data misfit estimated at all global points plus the original model. This process is done independently for each of the nine global ocean biogeochemical models and the final estimate is the average $pCO_2$.

directly measured or inferred from observations using carbonate system calculations with inputs of DIC and Total Alkalinity (TA). No data from GLODAPv2, BATS, or HOT are in the SOCAT database since $pCO_2$ is not directly measured at these sites, but instead is derived from DIC and TA. The cbsyst package (Hain et al., 2015) is used for carbonate system calculations. For decadal comparisons, timeframes are 1990s (1990–1999), 2000s (2000–2009) and 2010s (2010–2018).

The uncertainty in derived $pCO_2$ is dependent on the accuracy of the input measurements. For the modern ocean, cbsyst calculations are consistent with the constants of Lueker et al. (2000), and result in a 1.9% standard deviation in $pCO_2$ when DIC and TA uncertainties are 2.0 and 4.0 mol kg$^{-1}$, respectively. However, $pCO_2$ at

**Table 3**
*Validation Datasets, Accuracy of $pCO_2$, and Total Number of $1° \times 1°$ Grid Points Is Shown for Each Data Set*

| Data set | Accuracy ($\mu atm$) | Grid points | Reference |
|---|---|---|---|
| LDEO database version 2018[a] | $\pm 2.5 \ \mu atm$ | 16,161 | Takahashi et al. (2019) |
| GLODAPv2[b] | $>12 \ \mu atm$ at 400 $\mu atm$ | 5,976 | Olsen et al. (2019) |
| BATS[b] | $4 \ \mu atm$ at 400 $\mu atm$ | 246 | Bates (2007) |
| HOT[b] | $<7.6 \ \mu atm$ at 400 $\mu atm$ | 214 | Dore et al. (2009) |

[a]$pCO_2$ measured with $pCO_2$ equilibrator. [b]$pCO_2$ estimated from DIC and TA.

temperatures below 8°C may be underestimated using the Lueker et al. (2000) constants (Sulpis et al., 2020). For GLODAP, Bockmon and Dickson (2015) suggests an uncertainty of $5\frac{\mu mol}{kg}$ for DIC and TA, thus suggesting an uncertainty greater than 1.9%. Gregor et al. (2019) estimate the uncertainty of GLODAP $pCO_2$ to be >12 $\mu$atm at 400 $\mu$atm. Although the measurements have high uncertainty, given the sparsity of the SOCAT database, including GLODAP as a validation data set outweighs its omission, consistent with previous studies (Gregor et al., 2019; Gregor & Gruber, 2021). At BATS the uncertainty is about 4 $\mu$atm (Bates, 2007) while at HOT it is <7.6 $\mu$atm (Dore et al., 2009). LDEO $pCO_2$ has uncertainty of 2.5 $\mu$atm (Takahashi et al., 2019).

### 2.5. Regression Metrics

A suite of regression metrics are used to compare the predictions ($P$) to the observations ($O$; Stow et al., 2009). Metrics considered include correlation ($r$), bias, and root mean squared error (RMSE). Multiple metrics are considered in order to provide a thorough appraisal of each method. Metrics are displayed in a Taylor diagram (Taylor, 2001).

Pearson correlation coefficient ($r$) measures the tendency of the predicted and observations to vary together, bounded between, $-1 < r < 1$, with values near 1 indicating that they vary together and $-1$ indicating an inverse relationship. Correlation is also a measure of how well the phase is captured. Values near 1 and $-1$ indicate that the predictions and observations are perfectly in or out of phase, respectively. Intermediate values indicate a phase shift between the two signals, with values closer to zero indicating a larger phase shift between signals. The squared correlation $r^2$, or coefficient of determination, represents the variance explained by the regression. Correlation is defined as the covariance between predictions and observations divided by the product of their standard deviations, $r = \frac{\text{cov}(P,O)}{\sigma_P \sigma_O}$, $\sigma_P$ and $\sigma_O$ represent the standard deviation of the predictions and observations, respectively.

Bias, average absolute error (AAE), and RMSE each measure the size of discrepancies, with values near zero indicating a close match between predictions and observations. However, each metric has strengths and weaknesses. Bias is simply calculated as the long-term mean difference between predictions and observations ($bias = \overline{P} - \overline{O}$), where overbars represent the temporal mean. Positive and negative bias values indicate predictions that are generally overestimated and underestimated respectively. Thus, bias provides a measure of the direction of discrepancy. However, bias values falling close to zero can be misleading with significant positive offsets at one point in space or time canceling out significant negative offsets elsewhere. RMSE $= \sqrt{\overline{(P-O)^2}}$ measures of the magnitude of discrepancy, but squaring the misfit makes RMSE sensitive to outliers. Alternatively, AAE $= \overline{|P-O|}$ treats each misfit equally, but is a less commonly used metric. We report bias, AAE and RMSE since each one provides a different insight into the goodness-of-fit.

### 2.6. Area Coverage

The LDEO-HPD product covers 89.6% of the total ocean area, leaving out the Arctic and coastal zones. Before estimating the net carbon flux from observation-based products, we use the method of Fay et al. (2021) to fill spatial gaps in the $pCO_2$ product with climatology (Landschützer, Laruelle, et al., 2020) plus the global-mean trend. This fills in the 10.4% to create a global gap-free product. Climatological filling lowers global mean $pCO_2$ from 356 to 352 $\mu$atm in the final product. This climatological filling technique (Fay et al., 2021) was also applied to each observational data product to which we compare our results (Table 5).

### 2.7. Air-Sea $CO_2$ Flux

The air-sea $CO_2$ exchange was calculated using a bulk parameterization (Equation 2):

$$F_{CO_2} = k_w S_{CO_2}(1 - f_{ice})(pCO_2^{atm-moist} - pCO_2^{ocean}) \tag{2}$$

which parameterizes the air-sea $CO_2$ flux ($F_{CO_2}$) as a function of the gas transfer velocity ($k_w$), $CO_2$ solubility ($S_{CO_2}$), ice fraction ($f_{ice}$), and partial pressure of $CO_2$ in moist air ($pCO_2^{atm-moist}$) and surface ocean ($pCO_2^{ocean}$).

**Table 4**
*Wind Speed Products Used to Calculate $CO_2$ Flux*

| Wind speed product | Reference |
|---|---|
| CCMPv2.0[a] | Mears et al. (2019) |
| ERA5[b] | Hersbach et al. (2020) |
| JRA-55[c] | Harada et al. (2016) |

[a]Source: http://www.remss.com/measurements/ccmp/. [b]Source: https://www.ecmwf.int/en/forecasts/datasets/reanalysis-datasets/era5. [c]Source: https://jra.kishou.go.jp/JRA-55/.

Solubility is calculated following Weiss (1974) and partial pressure of moist air ($pCO_2^{atm-moist}$) is calculated following Equation 3,

$$pCO_2^{atm-moist} = xCO_2(P_{atm} - pH_2O) \tag{3}$$

where $xCO_2$ is the dry air mixing ratio of atmospheric $CO_2$, $P_{atm}$ is the total atmospheric pressure, and $pH_2O$ is the saturation vapor pressure (Dickson et al., 2007). We use the Wanninkhof (1992) formulation for the gas transfer velocity (Equation 4):

$$k_w = k_{w,scaled} u^2 \left( \frac{Sc}{660} \right)^{-0.5} \tag{4}$$

which parameterizes $k_w$ as a function of wind speed squared ($u^2$) and the Schmidt number ($Sc$). $k_w$ is scaled by a factor of $k_{w,scaled}$ for each wind product to match the invasion of bomb $^{14}C$ ($343 \pm 40 \times 10^{26}$ atoms $^{14}C$ as of 1994, see Sweeney et al., 2007) (Fay et al., 2021). Three wind products were used (Table 4). Flux was calculated separately for each wind product and then averaged to create the final best estimate.

$pCO_2$ measured in situ and compiled in the SOCAT database is set by the combination of the anthropogenic and natural background carbon cycles. Thus, the calculated flux is the net, or contemporary, flux ($F_{NET}$).

## 2.8. Estimating Anthropogenic Carbon Flux From the Net Flux

The net $CO_2$ flux is the sum of an anthropogenic and a natural component ($F_{NET} = F_{NAT} + F_{ANT}$). Surface ocean $pCO_2$ quantifies $F_{NET}$, while interior ocean data quantify $F_{ANT}$. Closure terms are required to compare these independent quantifications of the ocean carbon sink.

The dominant net air-sea flux due to the natural carbon cycle is the slow outgassing of riverine carbon by the ocean (Aumont et al., 2001). The community's estimate of the net riverine-induced carbon outgassing ($F_{RIV}$) is still evolving. Here we use an average of three estimates representing the spread of the available approaches: a geochemical budgeting perspective ($+0.45 \pm 0.18$ PgC/yr; Jacobson et al., 2007), a meridional heat constraint approach ($+0.78 \pm 0.41$ PgC/yr; Resplandy et al., 2018), and a process-based ocean model ($+0.23$ PgC/yr; Lacroix et al., 2020). Since no uncertainty is presented for the Lacroix et al. (2020) estimate, we assume a 50% $1\sigma$ uncertainty, which is consistent with the relative magnitude of uncertainty for the other two estimates. Combining these three estimates, we derive an estimate of carbon efflux due to river inputs to the ocean in the observation-based product flux estimates of $+0.49 \pm 0.26$ PgC/yr. This $F_{RIV} \approx F_{NAT}$ will be removed from $F_{NET}$ estimates from HPD and other products to arrive at $F_{ANT}$.

Anthropogenic carbon accumulation can be estimated from interior ocean observations, for which a global survey is completed approximately once per decade, and thus this component is estimated over a defined time period. Gruber, Clement, et al. (2019) find $F_{ANT}$ at $-2.6 \pm 0.3$ PgC/yr for 1994–2007. A changing ocean circulation may have modified $F_{NAT}$ over 1994–2007 through a non-steady state outgassing flux of natural carbon. Thus, a natural non-steady state flux ($F_{NAT,NS}$) has been proposed (Gruber, Clement, et al., 2019), that is, $F_{NAT} = F_{NAT,NS} + F_{RIV}$. Applying the transient steady state assumption to $F_{NET}$ from one observation-based product (Landschützer et al., 2016), Gruber, Clement, et al. (2019) find $F_{NAT,NS} = +0.38$ PgC/yr. However, the transient steady state assumption is known to hold when atmospheric carbon accumulation is exponential, and this has not been the case in recent decades (Raupach et al., 2014; Ridge & McKinley, 2020). This estimate of $F_{NAT,NS}$ is likely an upper bound. Nevertheless, we follow Gruber, Clement, et al. (2019) and adjust their $F_{ANT}$ estimate by this amount.

Adjusting the $F_{ANT}$ estimate of Gruber, Clement, et al. (2019) leads to $F_{ANT} + F_{NAT,NS} = -2.2 \pm 0.3$ PgC/yr for 1994–2007. Earlier, for the IPCC AR4, Denman et al. (2007) synthesized multiple estimates from ocean and atmosphere tracer studies to estimate $F_{ANT} = -2.2 \pm 0.4$ PgC/yr for 1990–1999, and without any adjustment for $F_{NAT,NS}$. We compare estimates of $F_{NET}-F_{RIV}$ from LDEO-HPD and other observation-based products (see Section 2.9) to these estimates.

**Table 5**
*Observational Data Products for Comparison to These Results*

| Observation-based pCO$_2$ product | Reference |
| --- | --- |
| MPI-SOMFFN | Landschützer et al. (2014); Landschützer, Gruber, et al. (2020) |
| JENA-MLS | Rödenbeck et al. (2014) |
| CMEMS | Denvil-Sommer et al. (2019) |
| CSIR | Gregor et al. (2019) |

### 2.9. Observational-Based Products

We compare the pCO$_2$ error statistics and CO$_2$ flux estimates to four products that extrapolate from SOCAT data to global coverage using machine learning or other statistical modeling techniques (Table 5).

## 3. Results

With LDEO-HPD, an XGB algorithm estimates time-varying maps of model-data misfit, and these misfits are then used to adjust model fields to arrive at an estimate of the real-world pCO$_2$, from which CO$_2$ flux is then calculated. By identifying large-scale patterns of model mismatch with observations (Section 3.1), LDEO-HPD approach reconstructs real-world pCO$_2$ with greater fidelity than other recently published approaches (Section 3.2). After correcting for riverine outgassing, air-sea CO$_2$ flux estimates from LDEO-HPD are consistent with independent observations for both 1990–1999 and 1994–2007 (Section 3.3).

### 3.1. Model-Data Misfit

The 9-model, global-mean bias of 10 $\mu$atm in ocean pCO$_2$ (Figure 1d) can partially be attributed to neglecting to account for the water vapor correction when calculating the atmospheric pCO$_2$ that forces the model (Dickson et al., 2007). If the molar concentration of CO$_2$ is measured in dry air then, by standard protocol (Orr et al., 2017), the atmospheric partial pressure of CO$_2$ must be reduced by the vapor pressure of water (Equation 3). This is typically a small percentage correction, but still a change in the pressure field of only 3% changes the partial pressure of CO$_2$ by about 10 $\mu$atm. Thus, if the water vapor correction is ignored, the partial pressure of CO$_2$ in the atmosphere that the ocean model experiences will be too high and ocean pCO$_2$ will also be high. Of the nine models, three do not account for this correction, and the other six do (Friedlingstein et al., 2020). Hauck et al. (2020) illustrate through comparison to SOCAT data that these models have a significant high bias in pCO$_2$. In addition, they show that several models that do include the water vapor correction also have a high pCO$_2$ bias, but do not identify the source of this error. The mean pCO$_2$ bias of +10 $\mu$atm that we find (Figure 1d) is thus partially, but not fully, attributable to several models not applying the water vapor correction.

The model corrections solved for by the XGB algorithm has significant spatial structure, and thus is doing far more than just addressing a global-mean bias in the GOBM priors. This is illustrated for two of the nine models in Figure 4. There are distinct patterns and consistent seasonality in the required corrections. For the MPI model, in the Southern Ocean and North Pacific, pCO$_2$ is far too high in winter (JJA) and far to low in summer (DJF), thus the XGB algorithm imposes strong negative and positive corrections, respectively. In the North Atlantic, however, winter is too low and summer is too high, requiring the opposite sign of corrections. In the subtropics, MPI requires a strong negative correction. For CNRM, these patterns are different, with the whole of the winter hemisphere generally being slightly too low in pCO$_2$ and the majority of the summer hemisphere being too high in pCO$_2$, requiring modest positive and negative correction, respectively. Both models require a positive correction in the equatorial Pacific. Zonal-average misfits (Figure 4b) indicate that both of these models require the same sign and comparable magnitude seasonal correction in the extratropical Northern Hemisphere, while MPI requires much larger corrections in the Southern Ocean.
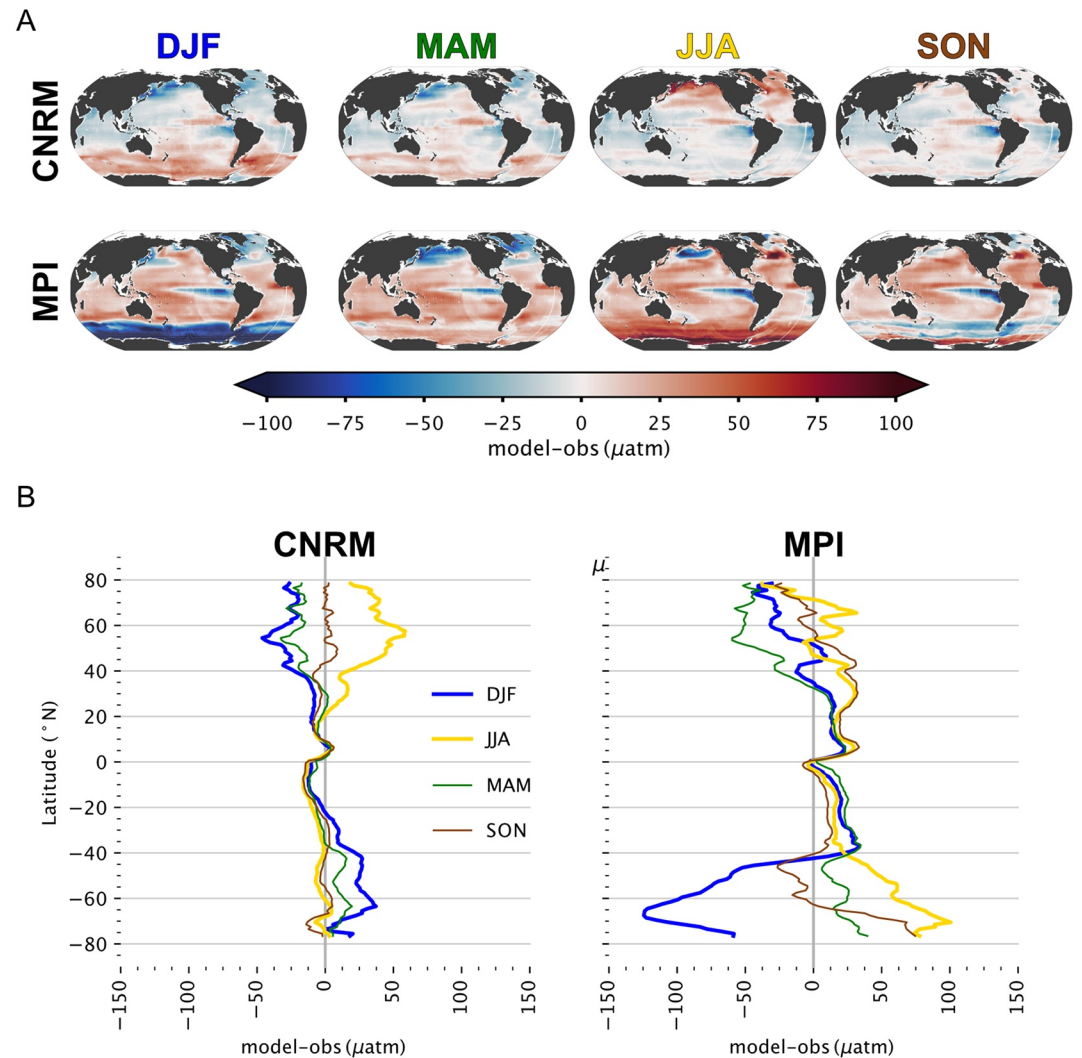
**Figure 4.** (a) Average pCO$_2$ misfit in the MPI and CNRM model for all years, December, January, and February (DJF); March, April and May (MAM); June, July, and August (JJA); and September October and November (SON). (b) Zonally average pCO$_2$ misfit in the MPI and CNRM models for DJF, JJA, MAM, and SON.

### 3.2. Evaluation of LDEO-HPD Against Independent Datasets

At the ocean timeseries sites at Bermuda (BATS) and Hawaii (HOT), LDEO-HPD compares quite favorably to the observations. The amplitude of seasonal and interannual variability in LDEO-HPD is as observed at HOT (Figures 5b and 6a) and slightly underestimated at BATS (Figures 5c and 6b), and the trends at both timeseries are well-represented. Compared to existing pCO$_2$ gap-filling methods, LDEO-HPD performs slightly better at BATS and HOT, with the lowest unbiased-RMSE relative to SOMFFN, MLS, and CMEMS (Figures 6a and 6b). Correlations are high at HOT and BATS because of the pronounced subtropical seasonality captured in the datasets. All these products are reliably able to capture subtropical seasonality (Gloege et al., 2021; Rödenbeck et al., 2015; Stamell et al., 2020).

LDEO and GLODAP are global observation datasets from intermittent ship transects. In these data, seasonality is less well-resolved, a fact that helps to explain the lower correlations of all products to the data relative to BATS and HOT. All the products show similar performance on LDEO observations, with all the products underestimating the variability (Figure 6c). For comparison to GLODAP, LDEO-HPD has a smaller unbiased-RMSE relative to other products (Figure 6d). LDEO-HPD and MLS capture the amplitude of variability in GLODAP equally well, and slightly better than SOMFFN and CMEMS.
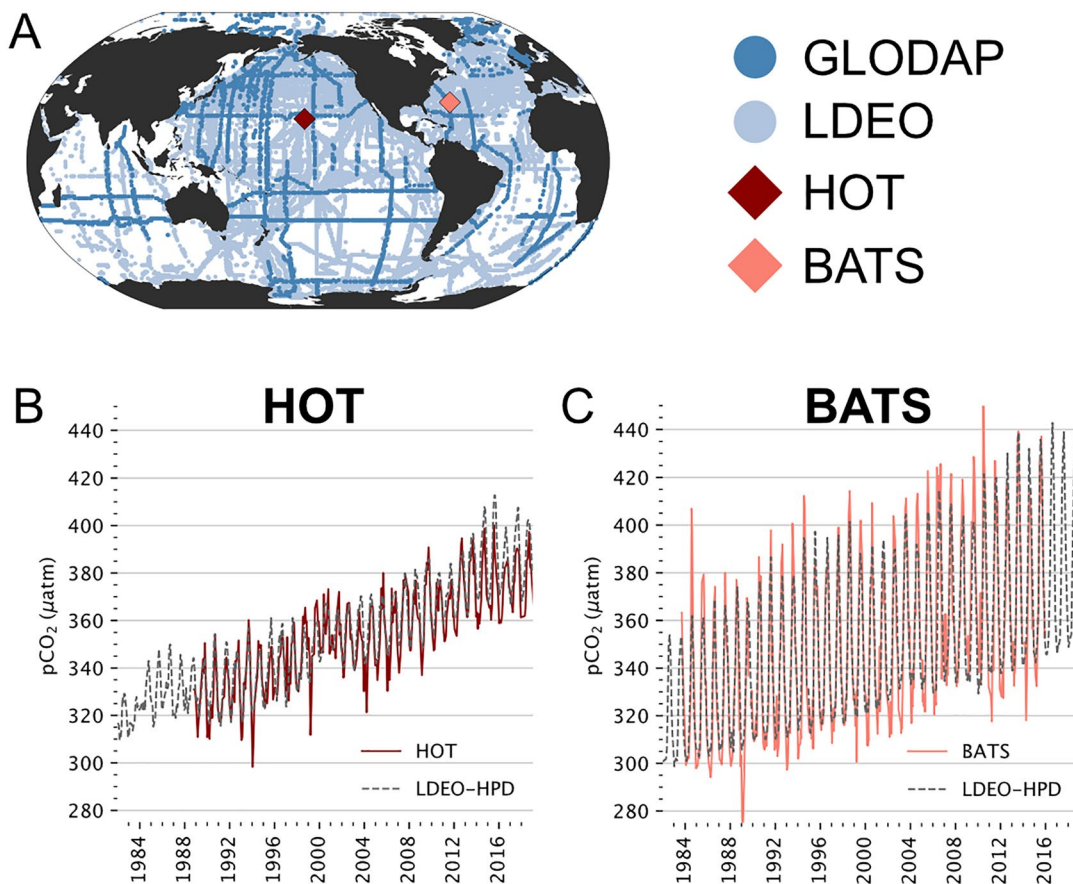
**Figure 5.** (a) Locations of independent datasets. Bermuda Atlantic Time-series Study (BATS) and Hawaii Ocean Time-series (HOT) are timeseries, while the GLODAP and Lamont-Doherty Earth Observatory are spatially varying. (b) Comparison of HOT with Lamont Doherty Earth Observatory-Hybrid Physics Data (LDEO-HPD) output. (c) comparison between BATS and LDEO-HPD output.

Over time, the skill of LDEO-HPD against independent observations of LDEO and GLODAP increases relative to the other methods. In the 1990s, the skill of all methods are indistinguishable (Figure 7, left). In the 2000s, comparison to GLODAP indicates that LDEO-HPD is slightly better than the others, though there is no distinction across the methods for LDEO (Figure 7, center). In the 2010s, LDEO-HPD clearly does the best job at capturing GLODAP, and is slightly improved against LDEO (Figure 7, right). Thus, we attribute the improved overall fit to independent observations (Figure 6) to the better fit in the later decades (Figure 7).

### 3.3. $CO_2$ Fluxes: 1982–2018

Mean $pCO_2$ and $CO_2$ flux from LDEO-HPD algorithm for 1982–2018 show well known features. Elevated $pCO_2$ is observed at the equator (Figure 8a), especially in the east-equatorial Pacific. This elevated $pCO_2$ is the result of upwelling of cold, carbon laden waters. The surface $pCO_2$ in this region is greater than the atmosphere, resulting in net $CO_2$ flux from the ocean to the atmosphere (Figure 8b).

Over time, the net global $CO_2$ flux has become increasingly negative (Table 6), that is, the ocean has become a greater net carbon sink over the recent decades as atmospheric $pCO_2$ has risen. Coastal filling (Section 2.6) increases uptake by nearly 0.1–0.2 PgC/yr, consistent with past estimates of globally integrated coastal uptake (Roobaert et al., 2019).

Applying the same to the calculation of air-sea $CO_2$ fluxes for all products (Section 2.7), and applying the $F_{RIV}$ correction, we find that fluxes estimated by LDEO-HPD are within the range of the other products for $F_{ant}$ (Figure 9a). Independent flux estimates based on interior data or atmospheric constraints also indicate consist-
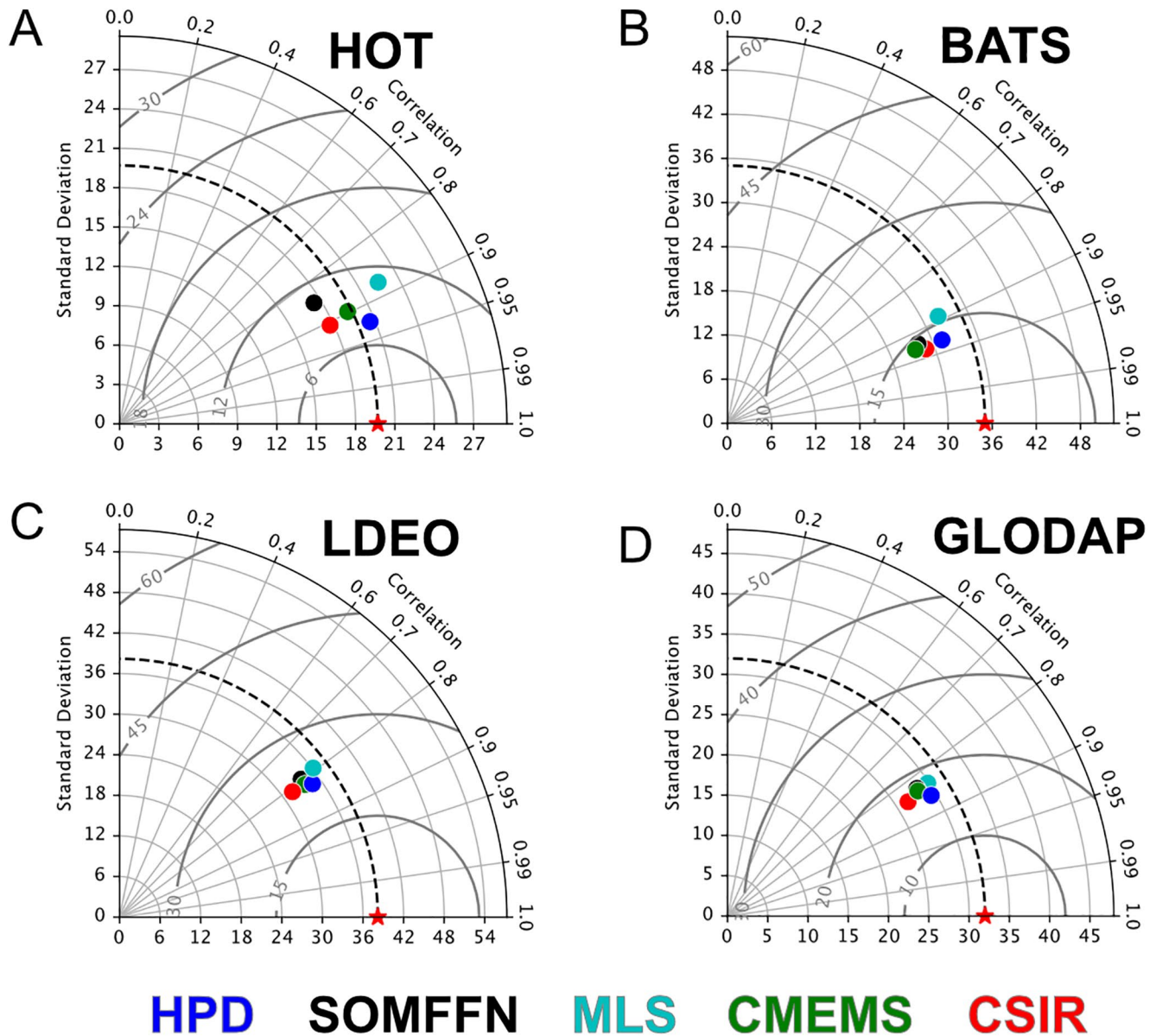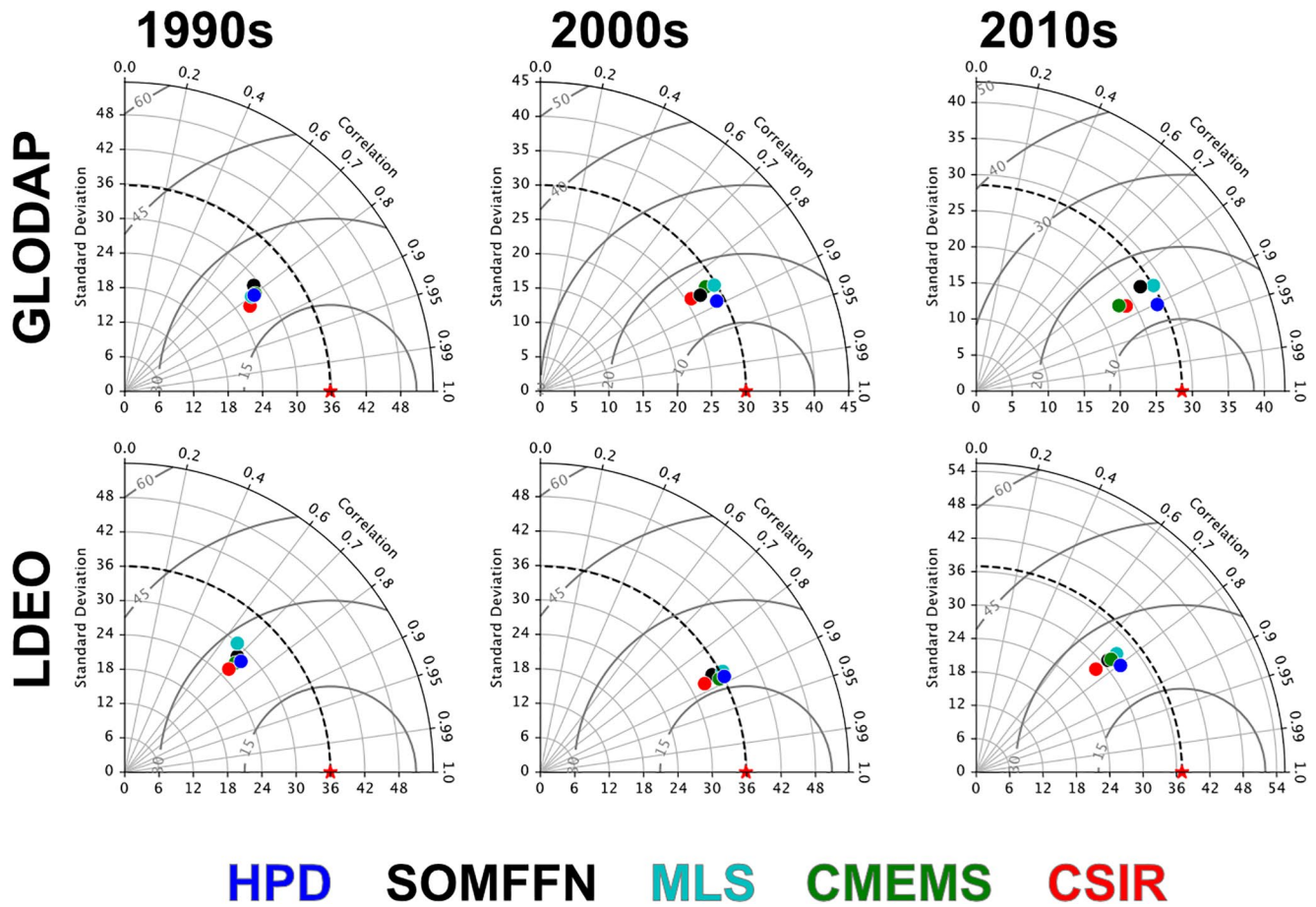
**Figure 6.** Taylor diagrams display the performance of published gap-filling techniques and Lamont Doherty Earth Observatory-Hybrid Physics Data (LDEO-HPD) product. Performance is evaluated at two timeseries: (a) Hawaii Ocean Time-series (HOT) and (b) Bermuda Atlantic Time-series Study (BATS); and two global datasets: (c) Lamont Doherty Earth Observatory and (d) GLODAP. Red star indicates standard deviation of each data set.

ency. Compared to $F_{ANT}$ for 1990–1999 (Denman et al., 2007) and $F_{ANT} + F_{NAT,NS}$ 1994–2007 (Gruber, Clement, et al., 2019), all products are within the uncertainty bounds (Figure 9b).

Improved comparison to independent data in LDEO-HPD is consistent with the reduced skewness of the target variable distribution (Figure 2). Reduced skewness should particularly improve predictions at the tails of the distribution, which in this case are the decades of the 1980s and 2010s. We do not have sufficient independent data to make comparisons in the 1980s, but HPD performs best of all methods in the 2010s (Figure 7, Table 7).

## 4. Discussion

We show that incorporating physical models into machine learning algorithms results in some improvement in predictions of surface ocean $pCO_2$. Using output from GOBMs as a prior guess allows us to reduce the skewness of the target variable distribution (Figure 2). Though GOBMs are imperfect representations of the real ocean (Hauck

**Figure 7.** Taylor diagrams display the performance of published gap-filling techniques and Lamont Doherty Earth Observatory-Hybrid Physics Data (LDEO-HPD) product. Performance is evaluated at two global data sets, Lamont Doherty Earth Observatory and GLODAP, using data from 1990 to 1999 (1990s), 2000–2009 (2000s), and 2010–2018 (2010s). Red star indicates standard deviation of each data set.
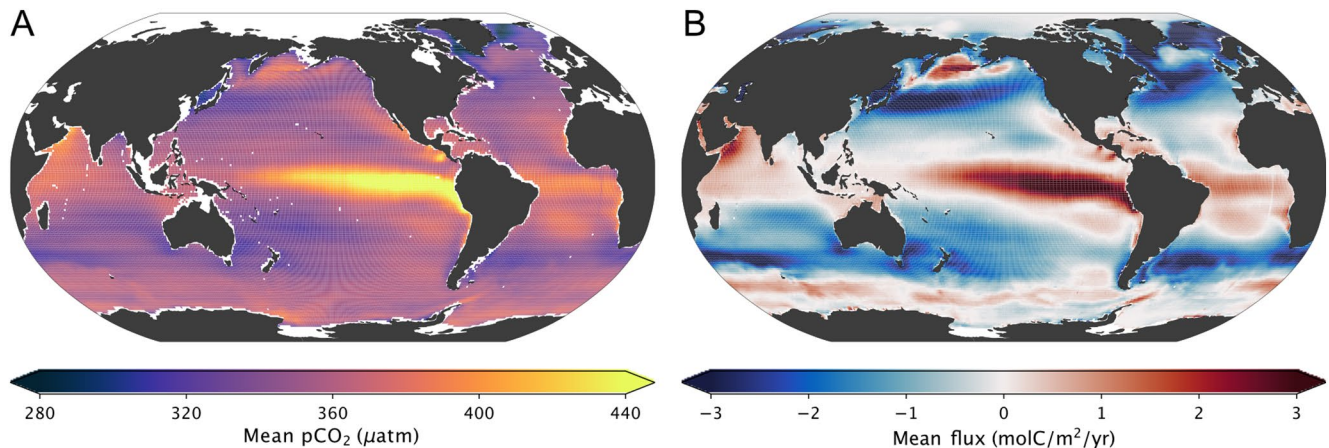


**Figure 8.** Mean (a) $pCO_2$ and (b) net $CO_2$ flux over 1982–2018 estimated from Lamont Doherty Earth Observatory-Hybrid Physics Data (LDEO-HPD). A spatially complete map of $CO_2$ flux is achieved by filling in gaps with a trend plus climatology.

**Table 6**
*Net $CO_2$ Flux ($F_{NET}$) (PgC/yr) From LDEO-HPD Across Decades Without Coastal Filling ("Unfilled") and Filled With the Climatology (Section 2.6)*

|  | Unfilled | Filled |
|---|---|---|
| 1982–1990 | −1.38 | −1.53 |
| 1990–2000 | −1.48 | −1.65 |
| 2000–2010 | −1.49 | −1.69 |
| 2010–2018 | −1.96 | −2.23 |

et al., 2020), this work illustrates that they can provide useful prior estimates of $pCO_2$ upon which data can improve using machine learning algorithms. By merging models and data, LDEO-HPD reduces error in estimates of $pCO_2$ (Figure 6), with the recent decades being the most improved (Figure 7, Table 7).

The LDEO-HPD approach of correcting GOBMs estimates the misfit between model output and observed $pCO_2$ at all points in space and time (Figure 3). The seasonality of model-data misfits (Figure 4) indicate that the LDEO-HPD is correcting for errors in model representation of seasonal physical and biogeochemical processes, such as mixed layer deepening and biological processes. These machine-learning derived maps of model-data misfit could be applied as a diagnostic of model performance to offer a larger-scale perspective that
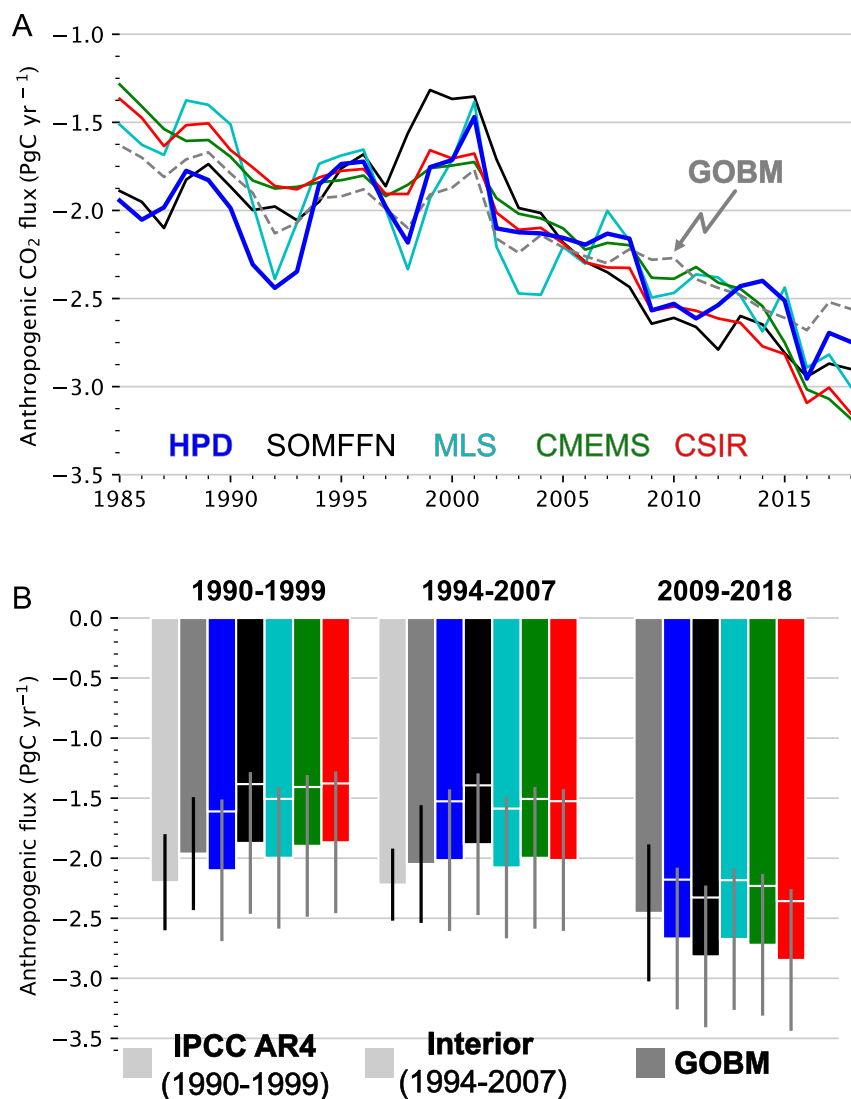


**Figure 9.** (a) Anthropogenic air-sea $CO_2$ exchange ($F_{ANT}$) for 1985–2018 from Lamont Doherty Earth Observatory-Hybrid Physics Data (LDEO-HPD) and four other products: SOMFFN, MLS, CMEMS, CSIR-ML6. Positive is to the atmosphere. Gray dash is the mean of the nine Global Ocean Biochemical Model (GOBM) priors, which are also the basis for the ocean sink estimate of the Global Carbon Budget 2020 (Friedlingstein et al., 2020). (b) Anthropogenic $CO_2$ flux for 1990–1999, 1994–2007, and 2009–2018. Light gray bar indicates IPCC AR4 or interior observation-based estimates with uncertainty. Dark gray bar is the mean of the nine GOBMs. Colored bars indicate observation-based estimates. The white line separates $F_{NET}$ from the products and $F_{RIV}$, estimated as the average of three estimates (0.49 PgC/yr), see Section 2.8. Thin lines indicate uncertainty; product uncertainty is estimated to be 0.6 PgC/yr, see Section 4.

**Table 7**
*RMSE at Independent Datasets Across Decades RMSE in Each Product Against GLODAP and LDEO Datasets Across Three Decades: 1990s, 2000s, and 2010s*

|  | 1990s | 2000s | 2010s |
|---|---|---|---|
| LDEO-HPD | 22.0 (27.6) | <u>13.8</u> (19.0) | <u>15.4</u> (<u>23.4</u>) |
| SOMFFN | 23.3 (28.2) | 15.4 (19.9) | 16.9 (26.0) |
| MLS | 22.2 (32.7) | 16.1 (25.5) | 17.7 (31.6) |
| CMEMS | 21.9 (<u>25.8</u>) | 16.2 (<u>18.6</u>) | 15.9 (24.8) |
| CSIR | <u>20.8</u> (28.4) | 15.6 (21.2) | 15.7 (27.9) |

*Note*. Underlined values indicate the product with the lowest RMSE. LDEO values are shown in parenthesis.

complements direct comparison to in situ data (Hauck et al., 2020). Model development could be supported with this approach to model-data comparison.

LDEO-HPD indicates an ocean carbon sink that is on the upper end of the suite of products for 1990–1999, but at the lower end of the suite of products for 2009–2018 (Figure 9). This may be due to reduced skewness in the target variable in this approach. The observed $pCO_2$ distribution is centered somewhere between the 1990s and 2000s (Figures 2a and 2b). The $pCO_2$ observed in the 1980s and 1990s is lower than this long-term mean $pCO_2$ due to the long-term increase of surface ocean $pCO_2$ in response to atmospheric $pCO_2$ growth, similarly the $pCO_2$ of the 2000s and 2010s is higher. This has the potential to skew $pCO_2$ predictions from machine learning algorithms in the 1980s and 1990s slightly high, and in the 2000s and 2010s slightly low, in both cases toward the mean of the overall distribution. This effect could help to explain the positive $pCO_2$ bias between 1982 and 1993 in the reconstruction of Gregor et al. (2019). Since LDEO-HPD has a reduced skewness in its target variable (Figures 2c and 2d), this potential issue should be minimized. Nevertheless, it is important to note that Landschützer et al. (2014) do not find a temporal trend in RMSE or bias for their product over 14 years (1998–2011). Yet, this is half the number of years as considered here, and any impact of skewness should increase with longer timeseries.

Machine learning algorithms are based on the assumption that training and testing data are independent and identically distributed and thus drawn from the same data generating distribution (Goodfellow et al., 2016). A tighter distribution, such as when the mean is not shifting over time, should be easier for a statistical algorithm to fit. By LDEO-HPD fitting model-data misfit, the skewness of the target variable distribution is largely eliminated (Figures 2c and 2d). This reduction of skewness in LDEO-HPD (Figures 2a and 2b) is consistent with both the improved fit to independent $pCO_2$ observations, particularly in later years (Figure 7, Table 7) and with the slightly larger ocean carbon sink LDEO-HPD in the 1990s and the slightly smaller sink since 2009 (Figure 9).

Here, we provide hints that temporal skewness of the $pCO_2$ training data may be impacting the current suite of observation-based products and LDEO-HPD's comparison to them. However, we do not prove this a significant issue. A comprehensive assessment is clearly needed, particularly in light of the strongly increasing product-based ocean carbon sink estimates since 2005 that increasingly diverge from models (Friedlingstein et al., 2020; Hauck et al., 2020) If the ocean sink is, in fact, growing much more rapidly than currently estimated by the Global Carbon Budget, this would imply much smaller trends must be occurring in carbon uptake by the land biosphere. A comprehensive study of this issue for the ocean products could be facilitated by tools such as the Large Ensemble Testbed (Gloege et al., 2021).

The combination of data-based machine learning with specific physical constraints or with the physics embodied in dynamical models is an emerging concept for earth science applications (Karpatne et al., 2017; Reichstein et al., 2019). As in other efforts that have corrected dynamical models using observations (Watt-Meyer et al., 2021), we use GOBMs as a prior estimate of the surface ocean $pCO_2$ field, and then correct these fields with data. The fact that the distribution of the target variable is substantially tightened (Figures 2a and 2c) suggests that GOBMs bring valuable prior physical information to support a robust reconstruction. For example, where $pCO_2$ is high, such as in the equatorial Pacific, it is also high in the model; and thus model-data misfits are constrained in magnitude (Figure 2c). If the GOBMs did not provide a useful prior, that is, had little relationship to the observations, the spread of model-data misfit would be expected to be larger than of $pCO_2$ alone. Tightening the distribution of the target variable supports our improved machine learning based predictions (Figures 6 and 7).

Gregor et al. (2019) suggest we may have "hit a wall" in our ability to extrapolate sparse $pCO_2$ data to global coverage. Here, we illustrate that incorporating model output and addressing skewness of the target variable distribution allows some additional improvement in prediction skill. In addition, LDEO-HPD employs an XGB algorithm, which is also found to be promising by Gregor et al. (2019). Stamell et al. (2020) showed the XGB algorithm performs slightly better in $pCO_2$ extrapolation than neural network or random forest algorithms. XGB's strength is its self-correcting nature in which each additional tree improves upon errors made in the previous.

For 2009–2018, the Global Carbon Budget 2020 (Friedlingstein et al., 2020) indicates an ocean anthropogenic sink ($F_{ANT}$) of $-2.5 \pm 0.6$ PgC/yr (Figure 9). LDEO-HPD indicates a similar flux, $-2.6 \pm 0.6$ PgC/yr ($F_{ANT} = F_{NET} -$

$F_{RIV}$). The standard deviation across the nine error-corrected GOBMs (0.1 PgC/yr), the uncertainty associated with $F_{RIV}$ (0.26 PgC/yr) and a 20% uncertainty from Wanninkhof (2014) associated with the gas transfer velocity (0.52 PgC/yr) are added in quadrature to produce the total uncertainty of LDEO-HPD. The other four products discussed here (Table 5) have mean uptake of $-2.6 \pm 0.6$ to $-2.8 \pm 0.6$ PgC/yr, using $F_{RIV} = +0.49 \pm 0.26$ PgC/yr to calculate $F_{ANT}$ from $F_{NET}$ for all and including a 20% uncertainty associated with the gas transfer velocity. Thus, all products are consistent with the Global Carbon Budget 2020.

It is important to note that our updated estimate of $F_{RIV}$ is lower than that used by the Global Carbon Budget 2020 (+0.61 PgC/yr), and by Hauck et al. (2020) (+0.78 PgC/yr), thus reducing the apparent model to observation-based product discrepancy that has been previously discussed (Friedlingstein et al., 2020). In addition, the harmonized flux calculation approach used here slightly reduces ocean uptake for some products (Fay et al., 2021). In summary, for 2009–2018, we find that all products fall within the uncertainties of the Global Carbon Budget 2020 for $F_{ANT}$, with LDEO-HPD on the lower end of the range and slightly closer to the Global Carbon Budget 2020 mean.

## 5. Conclusions

To reconstruct the real ocean's surface ocean $pCO_2$, LDEO-HPD rectifies output of nine GOBMs by learning the misfit from observed $pCO_2$ using an XGB algorithm and observed driver fields. LDEO-HPD improves prediction accuracy compared to other state-of-the-art $pCO_2$ data products, as indicated by improved fit to independent $pCO_2$ data. This suggests that GOBM output adds useful prior information to machine learning for this application. In addition, the globally and temporally complete misfits learned by the algorithm additionally have promise as a new diagnostic and visualization tool with which GOBM performance can be assessed.

Adding physical information, here by using GOBMs as a prior, and addressing temporal skewness in surface ocean $pCO_2$ distribution offer promising directions for continued improvement in the fidelity of machine-learning based reconstructions of the ocean carbon sink. The LDEO-HPD suggests a global ocean sink for anthropogenic carbon that is within the range of the suite of existing $pCO_2$ observation-based products, and that is in agreement with the Global Carbon Budget 2020 (Friedlingstein et al., 2020).

## Data Availability Statement

Analysis scripts and LDEO-HPD code is available at https://github.com/lgloege/LDEO-HPD and LDEO-HPD output is available at https://zenodo.org/record/4760205.

## References

Adcroft, A., Anderson, W., Balaji, V., Blanton, C., Bushuk, M., Dufour, C. O., et al. (2019). The GFDL global ocean and sea ice model om4. 0: Model description and simulation features. *Journal of Advances in Modeling Earth Systems*, *11*(10), 3167–3211. https://doi.org/10.1029/2019ms001726

Aumont, O., Éthé, C., Tagliabue, A., Bopp, L., & Gehlen, M. (2015). PISCES-v2: An ocean biogeochemical model for carbon and ecosystem studies. *Geoscientific Model Development Discussions*, *8*(2), 2465–2513. https://doi.org/10.5194/gmd-8-2465-2015

Aumont, O., Orr, J. C., Monfray, P., Ludwig, W., Amiotte Suchet, P., & Probst, J. L. (2001). Riverine-driven interhemispheric transport of carbon. *Global Biogeochem*, *15*, 393–405. https://doi.org/10.1029/1999gb001238

Bakker, D. C. E., Pfeil, B., Landa, C. S., Metzl, N., O'Brien, K. M., Olsen, A., et al. (2016). A multi-decade record of high-quality $f$co$_2$ data in version 3 of the surface ocean co$_2$ atlas (socat). *Earth System Science Data*, *8*(2), 383–413. https://doi.org/10.5194/essd-8-383-2016

Bates, N. R. (2007). Interannual variability of the oceanic co$_2$ sink in the subtropical gyre of the north atlantic ocean over the last 2 decades. *Journal of Geophysical Research*, *112*(C9), C09013. https://doi.org/10.1029/2006jc003759

Berthet, S., Séférian, R., Bricaud, C., Chevallier, M., Voldoire, A., & Ethé, C. (2019). Evaluation of an online grid-coarsening algorithm in a global eddy-admitting ocean biogeochemical model. *Journal of Advances in Modeling Earth Systems*, *11*(6), 1759–1783. https://doi.org/10.1029/2019ms001644

Bockmon, E. E., & Dickson, A. G. (2015). An inter-laboratory comparison assessing the quality of seawater carbon dioxide measurements. *Marine Chemistry*, *171*, 36–43. https://doi.org/10.1016/j.marchem.2015.02.002

Buitenhuis, E. T., Hashioka, T., & Quéré, C. L. (2013). Combined constraints on global ocean primary production using observations and models. *Global Biogeochemical Cycles*, *27*(3), 847–858. https://doi.org/10.1002/gbc.20074

Chen, T., & Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* (pp. 785–794). https://doi.org/10.1145/2939672.2939785

Conway, T. J., Tans, P. P., Waterman, L. S., Thoning, K. W., Kitzis, D. R., Masarie, K. A., & Zhang, N. (1994). Evidence for interannual variability of the carbon cycle from the national oceanic and atmospheric administration/climate monitoring and diagnostics laboratory global air sampling network. *Journal of Geophysical Research*, *99*(D11), 22831–22855. https://doi.org/10.1029/94jd01951

de Boyer Montégut, C., Madec, G., Fischer, A. S., Lazar, A., & Iudicone, D. (2004). Mixed layer depth over the global ocean: An examination of profile data and a profile-based climatology. *Journal of Geophysical Research*, *109*(C12), C12003. https://doi.org/10.1029/2004jc002378

Denman, K. L., Brasseur, G. P., Chidthaisong, A., Ciais, P., Cox, P. M., Dickinson, R. E., & Steffen, W. (2007). Couplings between changes in the climate system and biogeochemistry. In *Climate change 2007: The physical science basis* (pp. 499–588). Cambridge University Press.

Denvil-Sommer, A., Gehlen, M., Vrac, M., & Mejia, C. (2019). Lsce-ffnn-v1: A two-step neural network model for the reconstruction of Surface Ocean $pCO_2$ over the Global Ocean. *Geoscientific Model Development*, *12*(5), 2091–2105. https://doi.org/10.5194/gmd-12-2091-2019

Dickson, A. G., Sabine, C. L., & Christian, J. R. (2007). Guide to best practices for ocean $CO_2$ measurements. In *PICES Special Publication 3* (pp. 191).

Doney, S. C., Lima, I., Feely, R. A., Glover, D. M., Lindsay, K., Mahowald, N., & Wanninkhof, R. (2009). Mechanisms governing interannual variability in upper-ocean inorganic carbon system and air–sea $CO_2$ fluxes: Physical climate and atmospheric dust. *Deep Sea Research Part II: Topical Studies in Oceanography*, *56*(8–10), 640–655. https://doi.org/10.1016/j.dsr2.2008.12.006

Dore, J. E., Lukas, R., Sadler, D. W., Church, M. J., & Karl, D. M. (2009). Physical and biogeochemical modulation of ocean acidification in the central north pacific. *Proceedings of the National Academy of Sciences*, *106*(30), 12235–12240. https://doi.org/10.1073/pnas.0906044106

Fay, A. R., Gregor, L., Landschützer, P., McKinley, G. A., Gruber, N., Gehlen, M., et al. (2021). Harmonization of global surface ocean $pCO_2$ mapped products and their flux calculations; an improved estimate of the ocean carbon sink. *Earth System Science Data Discussions*, 1–32. https://doi.org/10.5194/essd-2021-16

Fay, A. R., & McKinley, G. A. (2013). Global trends in surface ocean $pCO_2$ from in situ data. *Global Biogeochemical Cycles*, *27*, 541–557. https://doi.org/10.1002/gbc.20051

Fay, A. R., & McKinley, G. A. (2017). Correlations of surface ocean $pCO_2$ to satellite chlorophyll on monthly to interannual timescales. *Global Biogeochemical Cycles*, *31*(3), 436–455. https://doi.org/10.1002/2016GB005563

Fay, A. R., & McKinley, G. A. (2021). Observed regional fluxes to constrain modeled estimates of the ocean carbon sink. *Geophysical Research Letters*, *48*, e2021GL095325. https://doi.org/10.1029/2021gl095325

Friedlingstein, P., Jones, M. W., O'Sullivan, M., Andrew, R. M., Hauck, J., Peters, G. P., et al. (2019). Global carbon budget 2019. *Earth System Science Data*, *11*(4), 1783–1838. https://doi.org/10.5194/essd-11-1783-2019

Friedlingstein, P., O'Sullivan, M., Jones, M. W., Andrew, R. M., Hauck, J., Olsen, A., et al. (2020). Global carbon budget 2020. *Earth System Science Data*, *12*(4), 3269–3340.

Gade, K. (2010). A non-singular horizontal position representation. *Journal of Navigation*, *63*(3), 395–417. https://doi.org/10.1017/s0373463309990415

Gloege, L., McKinley, G. A., Landschützer, P., Fay, A. R., Frölicher, T. L., et al. (2021). Quantifying errors in observationally based estimates of ocean carbon sink variability. *Global Biogeochemical Cycles*, *35*(4), e2020GB006788. https://doi.org/10.1029/2020gb006788

Good, S. A., Martin, M. J., & Rayner, N. A. (2013). En4: Quality controlled ocean temperature and salinity profiles and monthly objective analyses with uncertainty estimates. *Journal of Geophysical Research: Oceans*, *118*(12), 6704–6716. https://doi.org/10.1002/2013jc009067

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press. Retrieved from http://www.deeplearningbook.org

Gregor, L., & Gruber, N. (2021). OceanSODA-ETHZ: A global gridded data set of the Surface Ocean carbonate system for seasonal to decadal studies of ocean acidification. *Earth System Science Data*, *13*, 777–808. https://doi.org/10.5194/essd-13-777-2021

Gregor, L., Kok, S., & Monteiro, P. (2017). Empirical methods for the estimation of Southern Ocean $CO_2$: Support vector and random forest regression. *Biogeosciences*, *14*(23), 5551–5569. https://doi.org/10.5194/bg-14-5551-2017

Gregor, L., Lebehot, A. D., Kok, S., & Scheel Monteiro, P. M. (2019). 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*(12), 5113–5136. https://doi.org/10.5194/gmd-12-5113-2019

Gruber, N., Clement, D., Carter, B. R., Feely, R. A., Van Heuven, S., Hoppema, M., et al. (2019). The oceanic sink for anthropogenic $CO_2$ from 1994 to 2007. *Science*, *363*(6432), 1193–1199. https://doi.org/10.1126/science.aau5153

Gruber, N., Landschützer, P., & Lovenduski, N. S. (2019). The variable Southern Ocean carbon sink. *Annual Review of Marine Science*, *11*, 159–186. https://doi.org/10.1146/annurev-marine-121916-063407

Hain, M. P., Sigman, D. M., Higgins, J. A., & Haug, G. H. (2015). The effects of secular calcium and magnesium concentration changes on the thermodynamics of seawater acid/base chemistry: Implications for eocene and cretaceous ocean carbon chemistry and buffering. *Global Biogeochemical Cycles*, *29*(5), 517–533. https://doi.org/10.1002/2014gb004986

Harada, Y., Kamahori, H., Kobayashi, C., Endo, H., Kobayashi, S., Ota, Y., et al. (2016). The JRA-55 reanalysis: Representation of atmospheric circulation and climate variability. *Journal of the Meteorological Society of Japan. Series II*, *94*(3), 269–302. https://doi.org/10.2151/jmsj.2016-015

Hauck, J., Zeising, M., Le Quéré, C., Gruber, N., Bakker, D. C. E., Bopp, L., & Séférian, R. (2020). Consistency and challenges in the ocean carbon sink estimate for the Global Carbon Budget. *Frontiers in Marine Science*, *7*, 3167. https://doi.org/10.3389/fmars.2020.571720

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., et al. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 1999–2049. https://doi.org/10.1002/qj.3803

Jacobson, A. R., Mikaloff Fletcher, S. E., Gruber, N., Sarmiento, J. L., & Gloor, M. (2007). A joint atmosphere-ocean inversion for surface fluxes of carbon dioxide: 1. Methods and global-scale fluxes. *Global Biogeochemical Cycles*, *21*(1), 273. https://doi.org/10.1029/2005gb002556

Karpatne, A., Watkins, W., Read, J., & Kumar, V. (2017). *Physics-guided neural networks (pgnn): An application in lake temperature modeling*. arXiv preprint arXiv:1710.11431.

Lacroix, F., Ilyina, T., & Hartmann, J. (2020). Oceanic $CO_2$ outgassing and biological production hotspots induced by pre-industrial river loads of nutrients and carbon in a global modeling approach. *Biogeosciences*, *17*(1), 55–88. https://doi.org/10.5194/bg-17-55-2020

Landschützer, P., Gruber, N., Bakker, D., & Schuster, U. (2014). Recent variability of the global ocean carbon sink. *Global Biogeochemical Cycles*, *28*(9), 927–949. https://doi.org/10.1002/2014gb004853

Landschützer, P., Gruber, N., & Bakker, D. C. (2016). Decadal variations and trends of the global ocean carbon sink. *Global Biogeochemical Cycles*, *30*(10), 1396–1417. https://doi.org/10.1002/2015gb005359

Landschützer, P., Gruber, N., & Bakker, D. C. (2020). *An observation-based global monthly gridded sea surface pCO2 product from 1982 onward and its monthly climatology (NCEI Accession 0160558)*. Version 5.5. NOAA National Centers for Environmental Information. https://doi.org/10.7289/V5Z899N6

Landschützer, P., Laruelle, G. G., Roobaert, A., & Regnier, P. (2020). A uniform $pCO_2$ climatology combining open and coastal oceans. *Earth System Science Data*, 1–30.

Law, R. M., Ziehn, T., Matear, R. J., Lenton, A., Chamberlain, M. A., Stevens, L. E., et al. (2017). The carbon cycle in the australian community climate and earth system simulator (access-esm1)–part 1: Model description and pre-industrial simulation. *Geoscientific Model Development*, *10*(7), 2567–2590. https://doi.org/10.5194/gmd-10-2567-2017

Le Quéré, C., Raupach, M. R., Canadell, J. G., Marland, G., Bopp, L., Ciais, P., et al. (2009). Trends in the sources and sinks of carbon dioxide. *Nature Geoscience*, *2*(12), 831–836. https://doi.org/10.1038/ngeo689

Lueker, T. J., Dickson, A. G., & Keeling, C. D. (2000). Ocean pCO$_2$ calculated from dissolved inorganic carbon, alkalinity, and equations for k1 and k2: Validation based on laboratory measurements of co$_2$ in gas and seawater at equilibrium. *Marine Chemistry*, *70*(1–3), 105–119. https://doi.org/10.1016/s0304-4203(00)00022-0

Maritorena, S., d'Andon, O. H. F., Mangin, A., & Siegel, D. A. (2010). Merged satellite ocean color data products using a bio-optical model: Characteristics, benefits and issues. *Remote Sensing of Environment*, *114*(8), 1791–1804. https://doi.org/10.1016/j.rse.2010.04.002

McKinley, G. A., Fay, A. R., Eddebbar, Y. A., Gloege, L., & Lovenduski, N. S. (2020). External forcing explains recent decadal variability of the ocean carbon sink. *AGU Advances*, *1*(2), e2019AV000149. https://doi.org/10.1029/2019av000149

McKinley, G. A., Pilcher, D. J., Fay, A. R., Lindsay, K., Long, M. C., & Lovenduski, N. S. (2016). Timescales for detection of trends in the ocean carbon sink. *Nature*, *530*(7591), 469–472. https://doi.org/10.1038/nature16958

Mears, C. A., Scott, J., Wentz, F. J., Ricciardulli, L., Leidner, S. M., Hoffman, R., & Atlas, R. (2019). A near-real-time version of the cross-calibrated multiplatform (ccmp) ocean surface wind velocity data set. *Journal of Geophysical Research: Oceans*, *124*(10), 6997–7010. https://doi.org/10.1029/2019jc015367

Olsen, A., Lange, N., Key, R. M., Tanhua, T., Álvarez, M., Becker, S., et al. (2019). Glodapv2. 2019–an update of glodapv2. *Earth System Science Data*, *11*(3), 1437–1461. https://doi.org/10.5194/essd-11-1437-2019

Orr, J. C., Najjar, R. G., Aumont, O., Bopp, L., Bullister, J. L., Danabasoglu, G., et al. (2017). Biogeochemical protocols and diagnostics for the cmip6 ocean model intercomparison project (omip). *Geoscientific Model Development*, *10*(6), 2169–2199. https://doi.org/10.5194/gmd-10-2169-2017

Paulsen, H., Ilyina, T., Six, K. D., & Stemmler, I. (2017). Incorporating a prognostic representation of marine nitrogen fixers into the global ocean biogeochemical model hamocc. *Journal of Advances in Modeling Earth Systems*, *9*(1), 438–464. https://doi.org/10.1002/2016ms000737

Peters, G. P., Le Quéré, C., Andrew, R. M., Canadell, J. G., Friedlingstein, P., Ilyina, T., et al. (2017). Towards real-time verification of CO$_2$ emissions. *Nature Climate Change*, *7*(12), 848–850. https://doi.org/10.1038/s41558-017-0013-9

Raupach, M. R., Gloor, M., Sarmiento, J. L., Canadell, J. G., Frolicher, T. L., Gasser, T., et al. (2014). The declining uptake rate of atmospheric CO$_2$ by land and ocean sinks. *Biogeosciences*, *11*(13), 3453–3475. https://doi.org/10.5194/bg-11-3453-2014

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., & Prabhat (2019). Deep learning and process understanding for data-driven earth system science. *Nature*, *566*(7743), 195–204. https://doi.org/10.1038/s41586-019-0912-1

Resplandy, L., Keeling, R., Rödenbeck, C., Stephens, B., Khatiwala, S., Rodgers, K., et al. (2018). Revision of global carbon fluxes based on a reassessment of oceanic and riverine carbon transport. *Nature Geoscience*, *11*(7), 504–509. https://doi.org/10.1038/s41561-018-0151-3

Reynolds, R. W., Rayner, N. A., Smith, T. M., Stokes, D. C., & Wang, W. (2002). An improved in situ and satellite sst analysis for climate. *Journal of Climate*, *15*(13), 1609–1625. https://doi.org/10.1175/1520-0442(2002)015<1609:aiisas>2.0.co;2

Ridge, S., & McKinley, G. A. (2020). Ocean carbon uptake under aggressive emission mitigation. *Biogeosciences Discussions*. https://doi.org/10.5194/bg-2020-254

Rödenbeck, C., Bakker, D., Metzl, N., Olsen, A., Sabine, C., Cassar, N., et al. (2014). Interannual sea–air co 2 flux variability from an observation-driven ocean mixed-layer scheme. *Biogeosciences*, *11*(17), 4599–4613. https://doi.org/10.5194/bg-11-4599-2014

Rödenbeck, C., Bakker, D. C., Gruber, N., Iida, Y., Jacobson, A. R., Jones, S., et al. (2015). Data-based estimates of the ocean carbon sink variability–first results of the surface ocean pco$_2$ mapping intercomparison (socom). *Biogeosciences*, *12*, 7251–7278. https://doi.org/10.5194/bg-12-7251-2015

Roobaert, A., Laruelle, G. G., Landschützer, P., Gruber, N., Chou, L., & Regnier, P. (2019). The spatiotemporal dynamics of the sources and sinks of CO$_2$ in the global coastal ocean. *Global Biogeochemical Cycles*, *33*(12), 1693–1714. https://doi.org/10.1029/2019gb006239

Sabine, C. L., Hankin, S., Koyuk, H., Bakker, D. C. E., Pfeil, B., Olsen, A., et al. (2013). Surface ocean CO$_2$ atlas (socat) gridded data products. *Earth System Science Data*, *5*(1), 145–153. https://doi.org/10.5194/essd-5-145-2013

Sasse, T., McNeil, B., & Abramowitz, G. (2013). A novel method for diagnosing seasonal to inter-annual surface ocean carbon dynamics from bottle data using neural networks. *Biogeosciences*, *10*(6), 4319–4340. https://doi.org/10.5194/bg-10-4319-2013

Schwinger, J., Goris, N., Tjiputra, J. F., Kriest, I., Bentsen, M., Bethke, I., et al. (2016). Evaluation of noresm-oc (versions 1 and 1.2), the ocean carbon-cycle stand-alone configuration of the Norwegian earth system model (noresm1). *Geoscientific Model Development*, *9*, 2589–2622. https://doi.org/10.5194/gmd-9-2589-2016

Stamell, J., Rustagi, R. R., Gloege, L., & McKinley, G. A. (2020). Strengths and weaknesses of three machine learning methods for pcCO$_2$ interpolation. *Geoscientific Model Development Discussions*, 1–25

Stow, C. A., Jolliff, J., McGillicuddy, D. J., Jr, Doney, S. C., Allen, J. I., Friedrichs, M. A., & Wallhead, P. (2009). Skill assessment for coupled biological/physical models of marine systems. *Journal of Marine Systems*, *76*(1–2), 4–15. https://doi.org/10.1016/j.jmarsys.2008.03.011

Sulpis, O., Lauvset, S. K., & Hagens, M. (2020). Current estimates of k 1* and k 2* appear inconsistent with measured CO$_2$ system parameters in cold oceanic regions. *Ocean Science*, *16*(4), 847–862. https://doi.org/10.5194/os-16-847-2020

Sweeney, C., Gloor, E., Jacobson, A. R., Key, R. M., McKinley, G., Sarmiento, J. L., & Wanninkhof, R. (2007). Constraining global air-sea gas exchange for CO$_2$ with recent bomb 14c measurements. *Global Biogeochemical Cycles*, *21*(2), GB2015. https://doi.org/10.1029/2006gb002784

Takahashi, T., Sutherland, S. C., & Kozyr, A. (2019). *Global ocean surface water partial pressure of CO$_2$ database: Measurments performed during 1957–2018*. (version 2018).

Taylor, K. E. (2001). Summarizing multiple aspects of model performance in a single diagram. *Journal of Geophysical Research*, *106*(D7), 7183–7192. https://doi.org/10.1029/2000jd900719

Tjiputra, J. F., Olsen, A., Bopp, L., Lenton, A., Pfeil, B., Roy, T., et al. (2014). Long-term surface pco$_2$ trends from observations and models. *Tellus B: Chemical and Physical Meteorology*, *66*(1), 23083. https://doi.org/10.3402/tellusb.v66.23083

Wanninkhof, R. (1992). Relationship between wind speed and gas exchange over the ocean. *Journal of Geophysical Research*, *97*(C5), 7373–7382. https://doi.org/10.1029/92jc00188

Wanninkhof, R. (2014). Relationship between wind speed and gas exchange over the ocean revisited. *Limnology and Oceanography: Methods*, *12*(6), 351–362. https://doi.org/10.4319/lom.2014.12.351

Watt-Meyer, O., Brenowitz, N. D., Clark, S. K., Henn, B., Kwa, A., McGibbon, J., et al. (2021). Correcting weather and climate models by machine learning nudged historical simulations. *Geophysical Research Letters*, *48*(15), e2021GL092555. https://doi.org/10.1002/essoar.10505959.1

Weiss, R. (1974). Carbon dioxide in water and seawater: The solubility of a non-ideal gas. *Marine Chemistry*, *2*(3), 203–215.