Posted on 23 Nov 2022 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.1002/essoar.10507690.1 — This a preprint and has not been peer reviewed. Data may be preliminary.

Observed Regional Fluxes to Constrain Modeled Estimates of the Ocean Carbon Sink

Amanda R Fay[1] and Galen A McKinley[2]

[1]Lamont-Doherty Earth Observatory, Columbia University
[2]Lamont Doherty Earth Observatory of Columbia University

November 23, 2022

Abstract

We compare air-sea CO_2 exchange in an ensemble of global ocean hindcast models to a suite of observation-based products for 1990-2018. Individual products agree closely for regional fluxes, but individual models vary widely in their regional estimates. Despite their regional divergence, individual models estimate similar global mean fluxes, indicating that significant regional compensation occurs to balance in the global integral. Models diverge most strongly from the observed mean flux in the northern and southern subtropics, despite strong agreement in seasonality. In the Southern Ocean, models estimate a wide range of both mean and seasonality. The ensemble of observation-based products can be used to select the models that best represent the regionally-resolved mean state of air-sea CO_2 exchange. Three models are within 3σ of the observed estimates in all regions. With this selected model ensemble, the global mean flux is slightly reduced and its uncertainty is reduced by 35%.

Hosted file

`essoar.10507690.1.docx` available at https://authorea.com/users/529535/articles/602031-observed-regional-fluxes-to-constrain-modeled-estimates-of-the-ocean-carbon-sink

Observed Regional Fluxes to Constrain Modeled Estimates of the Ocean Carbon Sink

A. R. Fay[1], G. A. McKinley[1]

[1] Columbia University and Lamont Doherty Earth Observatory, Palisades NY, USA

Corresponding author: Amanda Fay (afay@ldeo.columbia.edu)

Key Points:

- Seven observation-based products produce consistent estimates of air-sea carbon flux mean and seasonality, both globally and regionally
- Only 3 of 9 ocean biogeochemical models agree (within 3) with observed regional mean fluxes; this metric is applied as a selection criteria
- From only the selected models, global mean flux and seasonality for 1990-2018 is more tightly constrained

Abstract

We compare air-sea CO_2 exchange in an ensemble of global ocean hindcast models to a suite of observation-based products for 1990-2018. Individual products agree closely for regional fluxes, but individual models vary widely in their regional estimates. Despite their regional divergence, individual models estimate similar global mean fluxes, indicating that significant regional compensation occurs to balance in the global integral. Models diverge most strongly from the observed mean flux in the northern and southern subtropics, despite strong agreement in seasonality. In the Southern Ocean, models estimate a wide range of both mean and seasonality. The ensemble of observation-based products can be used to select the models that best represent the regionally-resolved mean state of air-sea CO_2 exchange. Three models are within 3 of the observed estimates in all regions. With this selected model ensemble, the global mean flux is slightly reduced and its uncertainty is reduced by 35%.

Plain Language Summary

The ocean plays a crucial role in mitigating climate change by significantly modulating the growth of atmospheric carbon dioxide. Multiple ocean models that are regularly used to estimate the ocean's carbon uptake, all provide consistent estimates as to the magnitude and timing of the globally-integrated ocean carbon uptake. However, these same models indicate a wide range of regional patterns. Taking the observational perspective, it has recently been demonstrated that regional flux patterns can be constrained using existing data. Directly comparing the models to observation-based estimates at the regional scale, we find that only one-third of the models are consistent with the observed regional fluxes. Improvements to ocean model are needed to better represent ocean carbon uptake processes.

1 Introduction

The ocean plays a crucial role in mitigating climate change by significantly modulating the growth of atmospheric carbon dioxide (CO_2). Based on the Global Carbon Budget 2020 (Friedlingstein et al. 2020), the ocean took up 2.5 $\pm$ 0.6 PgC/yr, or 22% of the total anthropogenic emissions over the last decade (2009-2018). This sink estimate is based on an ensemble of ocean hindcast models. Though these models provide consistent estimates of the globally-integrated ocean sink and its long-term increase, they present significant regional, seasonal and interannual differences (Schuster et al. 2013; Wanninkhof et al. 2013; Mongwe et al. 2018). Models may also underestimate the interannual to decadal variability of the carbon sink (Landschützer et al. 2015; Le Quéré et al. 2018; Gruber et al. 2019a).

Observation-based products statistically interpolate pCO_2 observations to assess the global ocean carbon sink and its spatio-temporal variability. These methods build relationships between in-situ observations of pCO_2 and driver data via machine learning or diagnostic modeling. Since driver data have greater spatio-temporal coverage than in-situ pCO_2, the algorithms can be used to estimate pCO_2 at all points in space and time. Though in-situ pCO_2 observations only cover 1.5% of all 1^{o}x1^o degree, monthly locations across the globe (Bakker et al. 2016), the various approaches largely agree as to globally-integrated long-term mean flux and variability since the 1990s (Rödenbeck et al. 2015; Gregor et al. 2019; McKinley et al. 2020; Gloege et al. 2021b)

Recent work used a testbed of climate models to assess reconstruction skill of the MPI-SOMFFN product (Landschützer et al. 2014). MPI-SOMFFN can reconstruct surface ocean pCO_2 with small long-term mean bias globally, over large regions, and locally across most of the Northern Hemisphere (Gloege et al. 2021a). Seasonality can also be robustly reconstructed in most locations. Though Gloege et al. (2021a) only assess skill for MPI-SOMFFN, we show here that the other observation-based products cluster tightly around the regional MPI-SOMFFN results, and thus that they also are skillful. Despite low bias and robust seasonality, interannual to decadal variability can only be reconstructed with moderate skill by the observation-based products, with sparse sampling considered the key limiting factor (Rödenbeck et al. 2015; Stammel et al. 2020; Gloege et al. 2021a). Despite great interest in the interannual variability simulated by models (McKinley et al. 2017; Gruber et al. 2019a), it does not appear that the observation-based products are yet able to provide a robust constraint.

Strong agreement between globally-integrated temporal-mean fluxes from hindcast models and observation-based products has previously been demonstrated (McKinley et al. 2020; Friedlingstein et al. 2020; Hauck et al. 2020). McKinley et al. (2016) showed similarly strong agreement between MPI-SOMFFN and globally-integrated fluxes in the CMIP5 suite of climate models. Despite this global agreement, CMIP5 indicates significant spread in mean estimates for large ocean regions. This indicates compensation of regional biases in the global integration. Now that there are nearly as many individual observation-based products as models, and interpolation methods has been shown skillful (Gloege

et al. 2021a), we return to the question of regional skill in the hindcast models.

We focus on using the ensemble of observation-based products to assess model skill based on the 1990-2018 regional mean fluxes and seasonal amplitude and timing. Regional fluxes are important indicators of the underlying physical and biogeochemical processes that combine to create the ocean carbon sink (McKinley et al. 2017; Gruber et al. 2019a). Capturing mean fluxes is one indication that these processes are correctly simulated by the models, and should enhance confidence in estimates of interannual variability and of the future ocean carbon sink under a range of emission scenarios.

2 Methods

We evaluate ocean carbon fluxes from an ensemble of nine global ocean biogeochemical models and seven observation-based pCO_2 products for years 1990-2018. Flux is defined positive upward, i.e., CO_2 release from the ocean into the atmosphere is positive, and uptake by the ocean is negative. Evaluations are made for the global integral and regionally, based on biogeochemically-coherent regions (Fay & McKinley 2014; Figure 1).

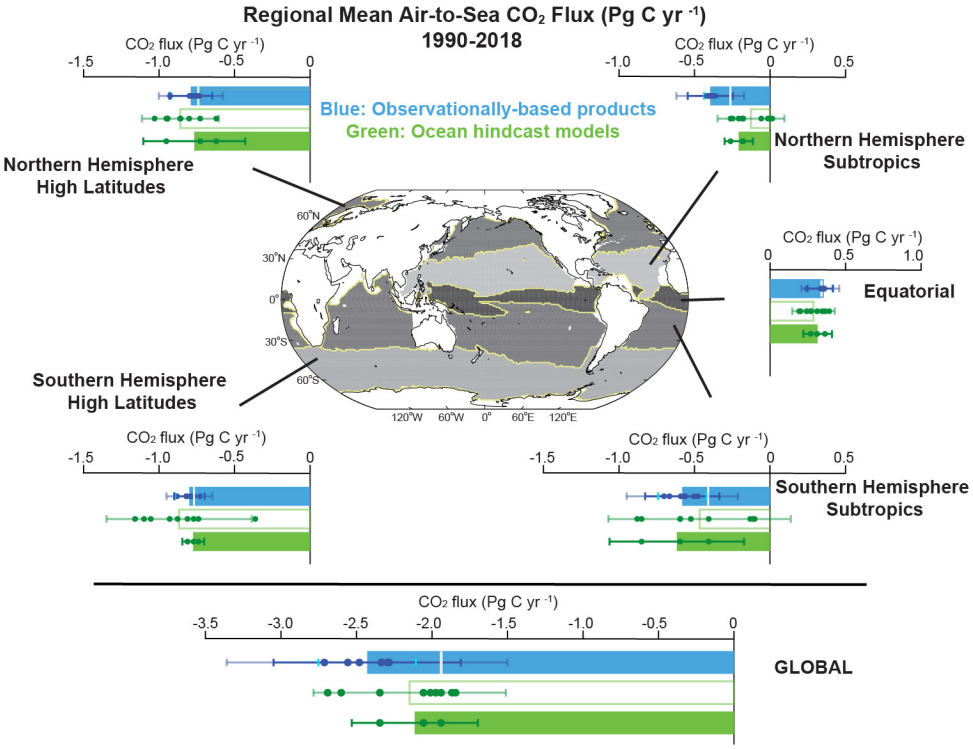

Figure 1 Global and regionally integrated air-sea CO_2 flux (PgC/yr) for ensemble mean of 7 observation-based products (blue) and 9 ocean hindcast models (green, solid and outlined) for 1990-2018 (Table S1). Green outlined bar is all

models; green solid bar is select models. White line on blue bar represents the global and regionally integrated mean without natural outgassing due to rivers. Map represents regional boundaries. Individual products/models represented by dots on bar. Green error bars represent 2 (95% confidence interval) on the mean for respective ensemble. Cyan error bars on blue bars represent the 2 (95% confidence interval) on the product ensemble without natural outgassing. Dark blue error bars represent 2 for the product ensemble with natural outgassing; opaque blue error bars show 3 bounds used in model criteria assessment.

2.1 Observation-based pCO_2 products

We utilize an ensemble of seven observation-based pCO_2 products (Table S1) and employ the SeaFlux package (Gregor & Fay 2021) to calculate fluxes from the products' estimated pCO_2. The SeaFlux package allows for consistent flux calculations for the product ensemble by applying area-coverage correction and appropriate scaling of the gas exchange coefficient (Fay et al. 2021). The observation-based pCO_2 product ensemble includes three neural network derived products (MPI-SOMFFN, CMEMS-FFNN, NIES-FNN), a mixed layer scheme product (JENA-MLS), a multiple linear regression (JMA-MLR), an eXtreme Gradient Boosting (XGB) method (LDEO-HPD), and a machine learning ensemble approach (CSIR-ML6).

2.2 Global Ocean Biogeochemical Models

Nine hindcast ocean models from the Global Carbon Budget (Friedlingstein et al. 2020) are utilized in this analysis (Table S1). They are all general ocean circulation models with coupled ocean biogeochemistry; additional details of the models can be found in Table A2, Friedlingstein et al. (2020). Here we utilize gridded fields of pCO_2 and air-sea CO_2 flux as submitted to the Global Carbon Budget. Models have been regridded to a common 1^ox1^o grid by their developers before submission.

As detailed in Hauck et al. (2020), model bias and drift in each model can be addressed by subtracting the constant climate simulation (Run B) from the contemporary simulation (Run A) which incorporates changes due to both increasing atmospheric CO_2 and climate. For the nine models included in the 2020 Global Carbon Budget, model biases range from -0.36 PgC/yr to 0.33 PgC/yr with a mean of -0.07 PgC/yr for years 1990-2018. Subtraction of Run B from Run A removes the bias and drift common to both simulations (Hauck et al. 2020). To apply this correction regionally, we area-weight the globally-integrated 1990-2018 flux bias and apply this to each region.

2.3 Regional analysis

For regional analysis we utilize the biomes of Fay & McKinley (2014) and combine individual biomes from each basin into five larger regions: Northern High Latitudes, Northern Subtropics, Equatorial, Southern Subtropics, and Southern High Latitudes (Figure 1). By comparing fluxes for these regions, we are further able to investigate the agreement of models and products. Our regional seasonal

analysis is conducted on 1990-2018 detrended monthly anomalies.

2.4 Natural outgassing due to carbon input from rivers

Fluxes calculated from the observation-based pCO_2 products are created from contemporary pCO_2 observations and thus represent the net flux. Hindcast models reflect the anthropogenic flux (Hauck et al. 2020, Eqns 1-2). In order to conduct model-product comparisons for the anthropogenic ocean carbon sink, an adjustment for the natural outgassing of carbon delivered to the ocean from land must be applied to the observation-based products. Quantitative understanding of this natural carbon outgassing is limited. We use an average of three estimates representing the spread of the available approaches: a geochemical budgeting perspective (0.45 ± 0.36 PgC/yr (2); Jacobson et al. 2007), a meridional heat constraint approach (0.78 ± 0.82 PgC/yr (2); Resplandy et al. 2018), and a process-based ocean model (0.23 PgC/yr; Lacroix et al. 2020). While there is no stated uncertainty on the Lacroix et al. (2020) estimate, we assume 100% uncertainty (2), comparable to relative magnitude in the other studies. The combined estimate, using the standard error of the mean to combine the uncertainties, is 0.49 ± 0.53 PgC/yr (2).

To apply the global riverine adjustment to our regional analysis, we utilize a gridded estimate from the model of Lacroix et al. (2020) and distribute the globally-integrated magnitude (0.49 ± 0.53 PgC/yr) proportionally (Table S2). Previous work (Friedlingstein et al. 2020; Hauck et al. 2020) uses the distribution reported by Aumont et al. (2001) in comparison of regional fluxes. The updated distribution of Lacroix et al. (2020) has a lesser percentage of efflux in the high latitude Southern Ocean and more in the subtropics (Table S2, S3). Uncertainty associated with the riverine efflux adjustment is summed in quadrature with that from the product estimates for each region, and shown separately (white bar, Figure 1). In all regions, the riverine flux adds significant uncertainty.

2.5 Statistical analysis

Statistical metrics presented within are based on the 1990-2018 mean of each model and product along with ensemble means for each group (Figure 1). We focus on a conservative uncertainty estimate of 2 bounds, which indicates 95% confidence that the mean of the ensemble falls within these bounds given the range present in the models/products. For the uncertainty presented on the product fluxes, we have summed in quadrature the uncertainty from the river estimate (globally, ± 0.53 PgC/yr) to that from the spread of the products: $\sqrt{(}$ ($_{\text{river}}{}^2$, $_{\text{product}}{}^2$)). We also note the mean and spread from the products themselves, without the addition from the river flux.

In addition to mean and 2 bounds we report the coefficient of variability (CoV) as a standardized measure of dispersion. It is expressed as a percentage and is defined as the ratio between the standard deviation and the mean: CoV = / * 100. This allows comparison between regions with large fluxes (and typically large standard deviations) such as the high latitude regions, with regions with

smaller fluxes (such as the subtropics and equatorial regions). Although the reported standard deviations may be small in regions with smaller flux, if the uncertainty is on the same order of magnitude as the mean flux itself, then CoV reflects this with a large percentage. With this statistic, relative uncertainties in regions with different mean fluxes can be more readily compared.

2.6 Model metric criteria

Given the tight regional estimates from the observation-based products and evidence that one of these products, MPI-SOMFFN, has low bias both globally and over large regions (Gloege et al. 2021a), we propose that the products can be used as a basis for selection of the best-performing models. Yet, we recognize that the products are not entirely independent because they are all based on the same sparse SOCAT pCO_2 data (Bakker et al. 2016). Thus, a liberal selection criteria is justified. Our criteria is to be within the 3 bounds (99.7% certainty) of the observation-based product ensemble mean in all five regions (Figure 1, 2). These 3 bounds include the regionally-variable uncertainty in natural outgassing due to rivers (Section 2.4).

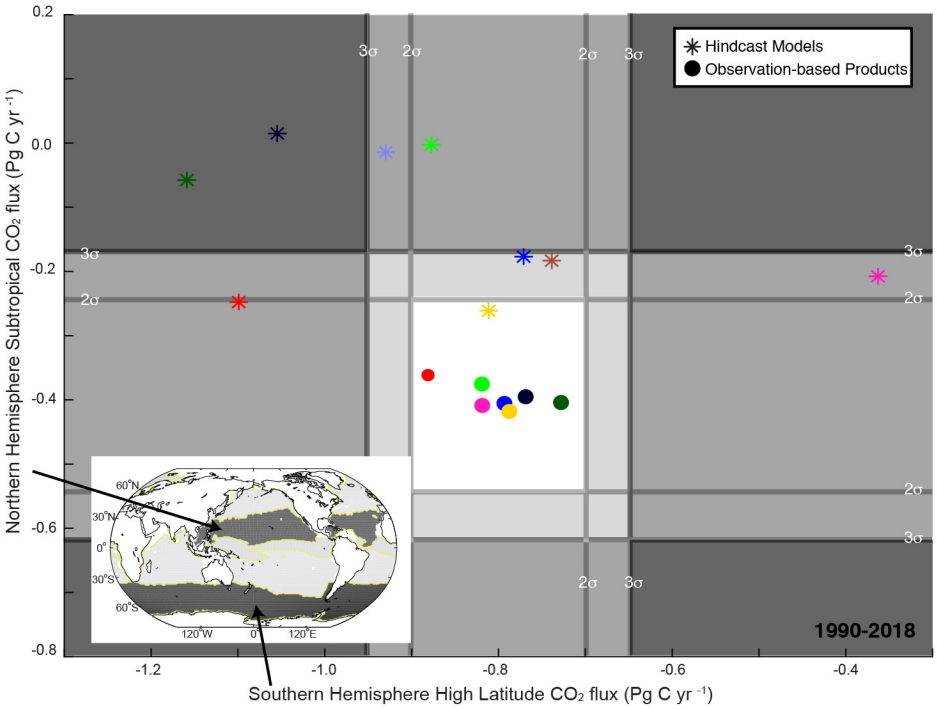

Figure 2 Regionally integrated air-sea CO_2 flux (PgC/yr) for ensemble of 9 ocean hindcast models (stars) and 7 observation-based products (dots) for the southern high latitude region (x-axis) and northern subtropical region (y-axis). Shading gradients represent 2 bounds (light gray) and 3 (darker gray). Inlaid

map indicates areas represented by the regions. Three models (stars) fall within the 3 bounds for these two regions and thus are selected. The uncertainty in natural outgassing is large in the subtropics (2 = 0.14PgC/yr), but small in the southern high latitudes (2 = 0.04 PgC/yr; Figure 1, cyan error bar vs blue error bar). This causes the 2 and 3 bounds to appear larger than based on the spread of the products (dots) in the y-axis scale.

3 Results

3.1 Globally-integrated Flux, 1990-2018

Globally integrated, model 1990-2018 mean fluxes are consistent with the observation-based products (Figure 1, bottom). Models estimate a mean anthropogenic flux of -2.15 ± 0.64 PgC/yr (2) and the products -2.43 ± 0.62 PgC/yr, after accounting for natural outgassing due to rivers. For the models, $\mathrm{CoV}_{\text{global models}} = 15\%$, while $\mathrm{CoV}_{\text{global products}} = 13\%$ (7% before inclusion of natural outgassing uncertainty).

This finding is consistent with previous comparisons of models and products (Devries et al. 2019; Friedlingstein et al. 2020; Hauck et al. 2020; McKinley et al. 2020) in that the observation-based products report a slightly greater ocean uptake flux than the models, but are consistent given uncertainties (Figure 1). With the incorporation of a larger product ensemble (seven products included here) and use of an updated river flux adjustment (0.49 PgC/yr), the major conclusions drawn for the globally-integrated flux remain unchanged from these previous studies. Both product and model ensemble globally-integrated fluxes are consistent with the independent interior observation-based estimate from Gruber et al. (2019b) who report an anthropogenic CO_2 flux of 2.6 ± 0.3 PgC/yr for 1994-2007. For the same time frame (1994-2007), the ensemble of models here has a globally-integrated anthropogenic flux of -2.00 ± 0.61 PgC/yr (2) while for the ensemble of observation-based products, it is -2.13 ± 0.90 PgC/yr (2).

3.2 Regional fluxes

Despite the modeled agreement for the global integral, models exhibit far less agreement with each other or with the observation-based products for regionally-integrated 1990-2018 mean fluxes. For each region, the spread of the model ensemble is larger than that of the product ensemble, ranging from 1 to nearly 5 times the magnitude. The 2 bound is always greater for the models than for the products (Table S4). Regionally, the CoV for the products is never greater than 21% (14% before inclusion of natural outgassing uncertainty), but it is always larger than this in the models (Table S5). The highest CoV regionally is in the northern subtropics ($\mathrm{CoV}_{\text{NHST models}} = 88\%$).

Modeled seasonal cycles agree to varying degrees with the observation-based products (Figure 3). In the subtropics, models are very consistent in timing and amplitude of seasonality when compared to the products. In the high northern latitudes, wintertime uptake is too large and summertime outgassing is

overestimated; and only one individual model captures the observed neutrality of the CO_2 flux in late winter. In the southern high latitudes, the products strongly agree as to the timing and magnitude of winter uptake and summer outgassing (Figure 3, bottom left). However, the models are widely spread in their estimates of both phase and amplitude. All but one model indicates a cycle magnitude that is consistent with the products, but several indicate the opposite phasing. Because these errors in the individual models cancel, the ensemble mean of all models (gray bold) is, to first-order, consistent with the products.

Interannual variability is not the focus of this analysis, but are included for completeness in the Supplementary Material.

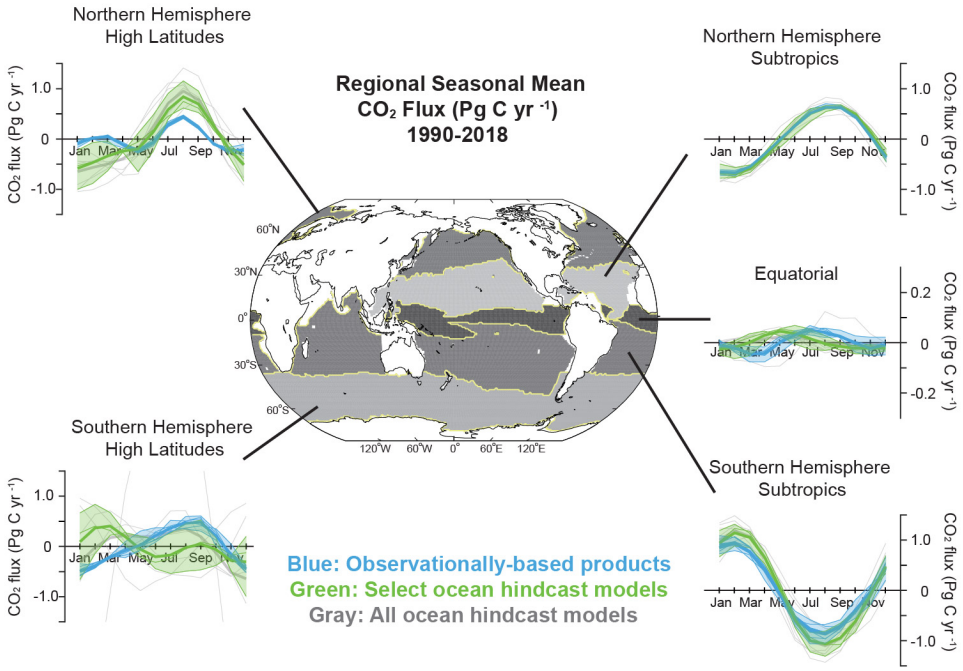

Figure 3 Regional air-sea CO_2 flux (PgC/yr) seasonal cycle for observation-based product ensemble (blue), ocean hindcast model ensemble (gray), and select models in green; mean of each ensemble shown in thick line of corresponding color. X-axis is month and y-axis is detrended monthly anomalies for each region. Shading represent 2 (95% confidence interval) on the seasonal mean of the model/product ensemble. All plots are on the same y-axis scale except for the equatorial.

3.3 Model selection

Despite model-product agreement for globally-integrated fluxes, we find models differ substantially from the products at the regional scale. The observation-based products are much more consistent across their ensemble. Further, a recent assessment of one of these products demonstrates low bias (Gloege et al. 2021a). Thus, we propose that the regional mean fluxes of the products can be used as a basis for selection of the best-performing models. Our criteria is that the model should estimate a mean flux that is within the 3 bounds (99.7% certainty) of the observation-based products in all five regions (Figure 1, 2).

Most models fall within this 3 criteria in the northern high latitude and equatorial regions. However, many models diverge from the observation-based products in the northern and southern subtropics, and in the southern high latitudes (Figure 1). The northern subtropics and southern high latitudes best demonstrate the application of the selection criteria (Figure 2). Only three models fall within the 3 criteria from the products in both of these regions. Two models are outside 3 in both regions (darkest gray in Figure 2), and 4 are outside 3 in either one of the other regions (medium gray in Figure 2). Were a 2 criteria used (white in Figure 2), all but one model would be excluded.

Comparing between all models and the selected models, the globally-integrated flux is minimally changed ($\mathrm{CoV}_{\text{global models}} = 15\%$; $\mathrm{CoV}_{\text{global select models}} = 10\%$); however, regional mean and seasonal differences are considerable (Figure 1 and 3). In all but one region (the northern high latitudes), the regional means from the selected models have a smaller spread (Table S5). In the equatorial and southern subtropics, selection reduces the CoV by roughly a third, while in the northern subtropics it is reduced from 88% to 23% and in the southern high latitudes from 28% to 5%.

Comparison of the root mean squared error (RMSE) between the ensemble-mean seasonal cycles of the observation-based products and of the select models indicates strong improvement in seasonal coherence (Figure 3, Table S6). The greatest improvement is seen in both the high latitude regions (RMSE reduced by 0.10 PgC/yr in the north and 0.28 PgC/yr in the south). Models with the most extreme seasonal amplitude in the southern high latitude seasonality did not meet selection criteria.

4 Discussion

In the ocean hindcast models, globally-integrated 1990-2018 mean fluxes are consistent with the ensemble of observation-based products (Figure 1, bottom). However, regional mean fluxes are far less consistent (McKinley et al. 2006; Schuster et al. 2013; McKinley et al. 2016; Hauck et al. 2020). Regional mean fluxes are first-order indicators of the physical and biogeochemical mechanisms of the ocean carbon sink (McKinley et al. 2017; Gruber et al. 2019b). Recent developments in statistical analysis of sparse pCO_2 data have led to observation-based products robustly represent global and regional air-sea CO_2 fluxes on long-term and seasonal timescales (Gloege et al. 2021a). We propose that an ensemble of these products can be used to assess the regional skill of ocean

hindcast models.

Compared to the models, the spread across the ensemble of observation-based products is much smaller in all regions, especially before the natural outgassing due to rivers is included (Figure 1, light blue ticks, Table S4). Models have a five-times larger 2 bounds in the northern subtropical region, before uncertainty due to natural outgassing from rivers is included. Yet, the northern subtropics is an area of strong agreement in seasonal cycle phasing and amplitude (Figure 3). Past seasonal comparisons to subtropical time series data, such as Bermuda Atlantic Time Series station, also show good model performance (Ullman et al. 2009; Schuster et al. 2013). Even though the models tightly capture the seasonality here, they all underestimate uptake. In the subtropics, seasonality is driven primarily by temperature variations (Takahashi et al. 2002), and the models are clearly able to replicate this effect (Figure 3). However, this class of models also tends to have low net primary production (Bennington et al. 2009, Galbraith et al. 2010, Long et al. 2013), and this may be the cause of the underestimation of carbon uptake.

The Southern Ocean is a region of particular interest for the ocean carbon sink (Lovenduski et al. 2008; Landschutzer et al. 2015; Fay et al. 2018, Bushinsky et al. 2019, Gloege et al. 2021a). The region is variable spatially and temporally, but only sparsely observed. Adding autonomous float observations from 2015 to present to two observation-based products, Bushinsky et al. (2019) find a Southern Ocean sink reduced by 0.1 to 0.35 PgC/yr for 2015-2017. If these results were to hold over the full period of this study, it would imply that the flux estimates from all current products should be reduced in magnitude. Were this the case, then, at most, one additional model would be selected using our criteria (Figure 2); the other 5 models all indicate uptake into the Southern Ocean that is larger than the current ensemble of products.

4.1 Uncertainty in natural outgassing of river carbon

The spatial pattern of the natural carbon outgassing due to rivers that has been applied to the observation-based product flux estimates has an impact on comparisons to models (Figure 1). Uncertainty in the global flux and 2 out of 5 regional mean fluxes is dominated not by the spread of estimates of individual products, but instead by uncertainty in the natural outgassing (Table S4). The globally-integrated magnitude of the natural outgassing is one component of this uncertainty. Several recent analyses have used 0.6-0.78 PgC/yr (DeVries et al. 2019; Gruber et al. 2019b; Friedlingstein et al. 2020; Hauck et al. 2020), based on one of, or the average of two estimates (Jacobson et al. 2007; Resplandy et al. 2018). However, these studies have not accounted for the uncertainty in this closure term. Our estimate includes the model-based estimate of LaCroix et al. (2020) and accounts for the uncertainty, $+0.49 \pm 0.53$ PgC/yr (Section 2.4). Though this mean estimate is smaller, inclusion of uncertainty in this study means that the result is inclusive of the previous ones. Also worth noting, were a larger mean outgassing applied, the estimated anthropogenic flux would increase in all regions, leading to fewer models meeting the selection criteria in

the northern subtropics (Figure 2).

In the southern high latitudes, we find a closer mean flux agreement between the ensemble of observation-based products and both the full suite of models and the selected models (Figure 1) as compared to a recent regional analysis (Hauck et al. 2020). This difference is attributable to the larger ensemble of observation-based products used here (7 instead of 3) and to a slightly different set of models. It is also due to our use of a spatial pattern of natural outgassing that is based on state-of-the-art ocean model simulations (LaCroix et al. 2020), in which a smaller percentage of the total natural outgassing occurs in the Southern Ocean than occurred in the earlier ocean model of Aumont et al. (2001) (Table S3). Efforts to reduce the large uncertainty in both the magnitude and spatial pattern of the natural outgassing are urgently needed to improve future comparisons of model and observation-based products.

4.2 Implications

In this study, we compare the ocean carbon sink for anthropogenic carbon estimated by hindcast ocean models to a newly-available suite of observation-based products. We find that 2/3 of the models are unable to replicate the observed regional partitioning of the mean 1990-2018 sink, despite capturing the globally-integrated flux. External forcing is a critical driver for the globally-integrated carbon flux (McKinley et al. 2020). However, the mechanisms of the ocean circulation and biogeochemistry set regional flux patterns (Lovenduski et al. 2008; Ullman et al. 2009; McKinley et al. 2017; Ridge and McKinley 2020). Thus, these comparisons may be more telling as to the skill of state-of-the-art ocean models than skill assessments based only on the global integral (Friedlingstein et al. 2020).

This study focuses on the results from hindcast models; similar codes are used in the coupled climate models with which the future climate is predicted. Under all scenarios of future emissions, the ocean carbon sink will play a critical role in modulating climate change (Randerson et al. 2015; Zickfeld et al. 2016; Schwinger & Tjiputra 2018; Ridge & McKinley 2021). This study indicates the need for model development efforts to improve the representation of the ocean circulation and biogeochemical processes that together determine the magnitude and spatial patterns of air-sea CO_2 fluxes. These efforts should improve our ability to predict the future state of the carbon cycle and its role in climate.

5 Conclusions

We present an assessment of global and regional air-sea CO_2 fluxes for ensembles of hindcast models and observation-based pCO_2 products. Seven observation-based estimates of air-sea carbon fluxes indicate a globally-integrated flux of -2.43 $\pm$ 0.85 PgC/yr for 1990-2018. Nine ocean hindcast models provide a similar estimate, -2.15 $\pm$ 0.64 PgC/yr. Regionally, individual observation-based products indicate very similar fluxes as the ensemble mean ($CoV_{\text{global products}}$ = 13%; $CoV_{\text{regional products}}$ range 6-21%). In contrast, the hindcast models demonstrate significant regional disagreement ($CoV_{\text{regional models}}$ range 15-88%)

despite their relatively tight global consistence ($CoV_{\text{global models}} = 15\%$). Seasonally, models agree with the products most closely in the subtropical regions; yet subtropical mean fluxes are inconsistent with the products in many models. Thus, we demonstrate that skill in representation of subtropical seasonality does not imply skill in representation of mean fluxes.

Given these comparisons, and prior evidence of skill in the observation-based products (Rödenbeck et al. 2015; Gloege et al. 2021a), we use the ensemble spread (3) of the products in each region as the basis to select the models that best represent the ocean carbon sink. Only three of the nine models meet this criteria in all five ocean regions. The selected model ensemble more tightly constrains global mean fluxes, and regional seasonality is also modestly improved. Since regional CO_2 fluxes are more dependent on modeled ocean circulation and biogeochemical mechanisms than globally-integrated fluxes, future predictions from models that robustly represent regional fluxes are likely to be more reliable.

Data Availability

Model output is made available by the Global Carbon Budget (https://www.globalcarbonproject.org/carbonbud
Observation-based products gridded pCO_2 is available from each reference individually and linked through the pySeaFlux python package used to calculate the observation-based product fluxes (available at https://github.com/lukegregor/SeaFlux; doi.org/10.5281/zenodo.5078404).

References

Aumont, O., Orr, J. C., Monfray, P., Ludwig, W., Amiotte-Suchet, P., & Probst, J. L. (2001). Riverine-driven interhemispheric transport of carbon. *Global Biogeochemical Cycles*, *15*(2), 393-405.

Aumont, O., Ethé, C., Tagliabue, A., Bopp, L., and Gehlen, M. (2015) PISCES-v2: an ocean biogeochemical model for carbon and ecosystem studies, Geosci. Model Dev., 8, 2465–2513, https://doi.org/10.5194/gmd-8-2465-2015.

Bakker, D. C. E., B. Pfeil, C. S. Landa, N. Metzl, K. M. O'Brien, A. Olsen, et al. (2016), A multi-decade record of high-quality f CO2 data in version 3 of the Surface Ocean CO2 Atlas (SOCAT), Earth Syst. Sci. Data, 8(2), 383–413, doi:10.5194/essd-8-383-2016

Bennington, V., McKinley, G. A., Dutkiewicz, S., & Ullman, D. (2009). What does chlorophyll variability tell us about export and air-sea CO2 flux variability in the North Atlantic?. *Global Biogeochemical Cycles*, *23*(3).

Berthet, S., Séférian, R., Bricaud, C., Chevallier, M., Voldoire, A., and Ethé, C. (2019) Evaluation of an Online Grid-Coarsening Algorithm in a Global Eddy-Admitting Ocean Biogeochemical Model, J. Adv. Model. Earth Sy., 11, 1759–1783, https://doi.org/10.1029/2019MS001644.

Buitenhuis, E. T., Rivkin, R. B., Sailley, S., & Le Quéré, C. (2010). Biogeochemical fluxes through microzooplankton. Global Biogeochemical Cycles, 24(4), https://doi.org/10.1029/2009GB003601

Bushinsky, S. M., Landschützer, P., R.denbeck, C., Gray, A. R., Baker, D., Mazloff, M. R., et al. (2019). Reassessing Southern Ocean air-sea CO2 flux estimates with the addition of biogeochemical float observations. Global Biogeochemical Cycles, 33(11), 1370–1388. https://doi.org/10.1029/2019GB006176

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 pCO2 over the global ocean. GeoscientificModel Development, 12(5), 2091–2105. ht tps://doi.org/10.5194/gmd-12-2091-2019

DeVries, T., Le Quéré, C., Andrews, O., Berthet, S., Hauck, J., Ilyina, T., et al. (2019). Decadal trends in the ocean carbon sink. Proceedings of the National Academy of Sciences, 116(24), 11646–11651.

Doney, S. C., Lima, I., Feely, R. A., Glover, D. M., Lindsay, K., Mahowald, N., Moore, J. K., & Wanninkhof, R. (2009). Mechanisms governing interannual variability in upper-ocean inorganic carbon system and air–sea CO2 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

Fay, A. R., & McKinley, G. A. (2014). Global open-ocean biomes: mean and temporal variability. Earth System Science Data, 6(2), 273–284. https://doi.org/10.5194/essd-6-273-2014

Fay, A. R., Lovenduski, N. S., McKinley, G. A., Munro, D. R., Sweeney, C., Gray, A. R., et al. (2018). Utilizing the Drake Passage Time-series to understand variability and change in subpolar Southern Ocean pCO 2. *Biogeosciences*, *15*(12), 3841-3855.

Fay, A. R., Gregor, L., Landschützer, P., McKinley, G. A., Gruber, N., Gehlen, M., et al. (2021). Harmonization of global surface ocean pCO2 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

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.

Galbraith, E. D., Gnanadesikan, A., Dunne, J. P., & Hiscock, M. R. (2010). Regional impacts of iron-light colimitation in a global biogeochemical model. *Biogeosciences*, 7(3), 1043-1064. https://doi.org/10.5194/bg-7-1043-2010

Gloege, L., McKinley, G. A., Landschützer, P., Fay, A. R., Frölicher, T. L., Fyfe, J. C et al. (2021a). Quantifying errors in observationally based estimates of ocean carbon sink variability. *Global Biogeochemical Cycles*, *35*(4), https://doi.org/10.1029/2020GB006788.

Gloege, L. and Yan, M.a and Zheng, T. and McKinley, G. A. (2021b) Improved quantification of ocean carbon uptake by using machine learning to merge global models and pCO2 data. Earth and Space Science Open Archive. 10.1002/essoar.10507164.1.

Gregor, L., Lebehot, A. D., Kok, S., & Monteiro, P. M. S. (2019). A comparative assessment of the uncertainties of global surface ocean CO2 estimates using a machine learning ensemble (CSIR-ML6 version 2019a) – have we hit the wall? Geoscientific Model Development Discussion, 12(12), 5113–5136. https://doi.org/10.5194/gmd-12-5113-2019

Gregor, L., Fay, A. R. (2021) SeaFlux data set: Air-sea CO2 fluxes for surface pCO2 data products using a standardised approach. Zenodo, doi.org/10.5281/zenodo.5078404.

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

Gruber, N., Clement, D., Carter, B. R., Feely, R. A., Van Heuven, S., Hoppema, M., et al. (2019b). The oceanic sink for anthropogenic CO2 from 1994 to 2007. Science, 363(6432), 1193–1199. https://doi.org/10.1126/science.aau5153

Hauck, J., Zeising, M., Le Quéré, C., Gruber, N., Bakker, D. C. E., Bopp, L., Chau, T. T. T., Gürses, Ö., Ilyina, T., Landschützer, P., Lenton, A., Resplandy, L., Rödenbeck, C., Schwinger, J., and Séférian, R. (2020) Consistency and Challenges in the Ocean Carbon Sink Estimate for the Global Carbon Budget, Front. Mar. Sci., 7, 1–33, https://doi.org/10.3389/fmars.2020.571720.

Iida, Y., Takatani, Y., Kojima, A. and Ishii, M.: Global trends of ocean CO2 sink and ocean acidification: an observation-based reconstruction of surface ocean inorganic carbon variables. (2020) Journal of Oceanography, pp.1-36, https://doi.org/10.1007/s10872-020-00571-5.

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), GB1019. https://doi.org/10.1029/2005GB002556

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., Haumann, F. A., Rödenbeck, C., Bakker, D. C., Van Heuven, S., et al. (2015). The reinvigoration of the Southern Ocean carbon sink. Science, 349(6253), 1221–1224. https://doi.org/10.1126/science.aab2620

Landschützer, P., Gruber, N., Bakker, D. C. E. (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://www.ncei.noaa.gov/access/ocean-carbon-data system/oceans/SPCO2_1982_present_ETH_SOM_FFN.html. Dataset. https://doi.org/10.7289/V5Z899N6.

Law, R. M., Ziehn, T., Matear, R. J., Lenton, A., Chamberlain, M. A., Stevens, L. E., Wang, Y.-P., Srbinovsky, J., Bi, D., Yan, H., and Vohralik, P. F. (2017) The carbon cycle in the Australian Community Climate and Earth System Simulator (ACCESS-ESM1) – Part 1: Model description and pre-industrial simulation, Geosci. Model Dev., 10, 2567–2590, https://doi.org/10.5194/gmd-10-2567-2017.

Le Quéré, C., Andrew, R. M., Friedlingstein, P., Sitch, S., Hauck, J., Pongratz, J., et al. (2018). Global carbon budget 2018. Earth System Science Data (Online), 10(4), 2141–2194

Liao, E., Resplandy, L., Liu, J., and Bowman, K. W. (2020) Amplification of the Ocean Carbon Sink During El Niños: Role of Poleward Ekman Transport and Influence on Atmospheric CO2, Global Biogeochem. Cy., 34, https://doi.org/10.1029/2020GB006574.

Long, M. C., K. Lindsay, S. Peacock, J. K. Moore, and S. C. Doney (2013), Twentieth-Century Oceanic Carbon Uptake and Storage in CESM1(BGC). *J. Clim.*, 26 (18), 6775-6800, 10.1175/JCLI-D-12-00184.1.

Lovenduski, N. S., Gruber, N., & Doney, S. C. (2008). Toward a mechanistic understanding of the decadal trends in the Southern Ocean carbon sink. *Global Biogeochemical Cycles*, **22**(3). https://doi.org/10.1029/2007GB003139

McKinley, G. A., Takahashi, T., Buitenhuis, E., Chai, F., Christian, J. R., Doney, S. C et al. (2006). North Pacific carbon cycle response to climate variability on seasonal to decadal timescales. *Journal of Geophysical Research: Oceans*, *111*(C7). https://doi.org/10.1029/2005JC003173

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

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., Fay, A. R., Lovenduski, N. S., & Pilcher, D. J. (2017). Natural variability and anthropogenic trends in the ocean carbon sink. Annual Review of Marine Science, 9(1), 125–150. https://doi.org/10.1146/annurev-marine-010816-060529

Mongwe, N., Vichi, M., & Monteiro, P. (2018). The seasonal cycle of pCO2 and CO2 fluxes in the Southern Ocean: diagnosing anomalies in CMIP5 earth system models. Biogeosciences, 15(9), 2851–2872. https://doi.org/10.5194/bg-15-2851-2018

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

Randerson, J. T., Lindsay, K., Munoz, E., Fu, W., Moore, J. K., Hoffman, F. M., et al. (2015). Multicentury changes in ocean and land contributions to the climate-carbon feedback. Global Biogeochemical Cycles, 29(6), 744–759. https://doi.org/10.1002/2014GB005079

Resplandy, L., Keeling, R. F., Rödenbeck, C., Stephens, B. B., Khatiwala, S., Rodgers, K. B., Long, M. C., Bopp, L., and Tans, P. P. (2018) Revision of global carbon fluxes based on a reassessment of oceanic and riverine carbon transport, Nat. Geosci., 11, 504–509, https://doi.org/10.1038/s41561-018-0151-3.

Ridge, S. M., & McKinley, G. A. (2021). Ocean carbon uptake under aggressive emission mitigation. *Biogeosciences*, *18*(8), 2711-2725. https://doi.org/10.5194/bg-18-2711-2021

Rödenbeck, C., Keeling, R. F., Bakker, D. C., Metzl, N., Olsen, A., Sabine, C., & Heimann, M. (2013). Global surface-ocean p (CO2) and sea-air CO2 flux variability from an observation-driven ocean mixed-layer scheme. Ocean Sci., 9, 193–216, https://doi.org/10.5194/os-9-193-2013, 2013.

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 pCO2 Mapping intercomparison (SOCOM). Biogeosciences, 12(23), 7251–7278. https://doi.org/10.5194/bg-12-7251-2015

Schuster, U., McKinley, G. A., Bates, N., Chevallier, F., Doney, S. C., Fay, A. R. et al. (2013). An assessment of the Atlantic and Arctic sea–air CO 2 fluxes, 1990–2009. *Biogeosciences*, *10*(1), 607-627. https://doi.org/10.5194/bg-10-607-2013

Schwinger, J., Goris, N., Tjiputra, J. F., Kriest, I., Bentsen, M., Bethke, I., Ilicak, M., Assmann, K. M., and Heinze, C. (2016) Evaluation of NorESM-OC (versions 1 and 1.2), the ocean carbon cycle stand-alone configuration of the Norwegian Earth System Model (NorESM1), Geosci. Model Dev., 9, 2589–2622, https://doi.org/10.5194/gmd-9-2589-2016.

Schwinger, J. & Tjiputra, J. (2018) Ocean Carbon Cycle Feedbacks Under Negative Emissions. *Geophysical Research Letters* **26**, 5289.

Stamell, J., R.R. Rustagi, L. Gloege, and G.A. McKinley (2020) Strengths and weaknesses of three Machine Learning methods for pCO2 interpolation, Geoscientific Model Development Discussions doi:10.5194/gmd-2020-311 (60).

Takahashi, T., Sutherland, S. C., Sweeney, C., Poisson, A., Metzl, N., Tilbrook, B., et al. (2002). Global sea–air CO2 flux based on climatological surface ocean pCO2, and seasonal biological and temperature effects. *Deep*

Sea Research Part II: Topical Studies in Oceanography, *49*(9-10), 1601-1622. https://doi.org/10.1016/S0967-0645(02)00003-6

Ullman, D. J., McKinley, G. A., Bennington, V., & Dutkiewicz, S. (2009). Trends in the North Atlantic carbon sink: 1992–2006. *Global Biogeochemical Cycles*, *23*(4). https://doi.org/10.1029/2008GB003383

Wanninkhof R, Park GH, Takahashi T, Sweeney C, Feely R, et al. (2013). Global ocean carbon uptake: magnitude, variability and trends. Biogeosciences 10:1983–2000

Zeng, J., Nojiri, Y., Landschützer, P., Telszewski, M. and Nakaoka, S.I. (2014) A global surface ocean fco2 climatology based on a feed-forward neural network. Journal of Atmospheric and Oceanic Technology, 31(8), pp.1838-1849, https://doi.org/ 10.1175/JTECH-D-13-00137.1.

Zickfeld, K., MacDougall, A. H. & Matthews, H. D. (2016) On the proportionality between global temperature change and cumulative CO2 emissions during periods of net negative CO2 emissions. *Environ Res Lett* **11**, 055006.

Geophysical Research letters

Supporting Information for

Observed Regional Fluxes to Constrain Modeled Estimates of the Ocean Carbon Sink

A. R. Fay[1], G. A. McKinley[1]

1 Columbia University and Lamont Doherty Earth Observatory, Palisades NY, USA

Contents of this file

Text S1.

Globally and in the equatorial and subtropical regions, the ensemble of the observation-based products and the ensemble of models are in good agreement with regard to the phasing of interannual variability, both prior to and after selection (Supplementary Figure 1). The high latitudes show limited agreement in either case. Selection does not significantly change interannual correlations (Supplementary Table 7).

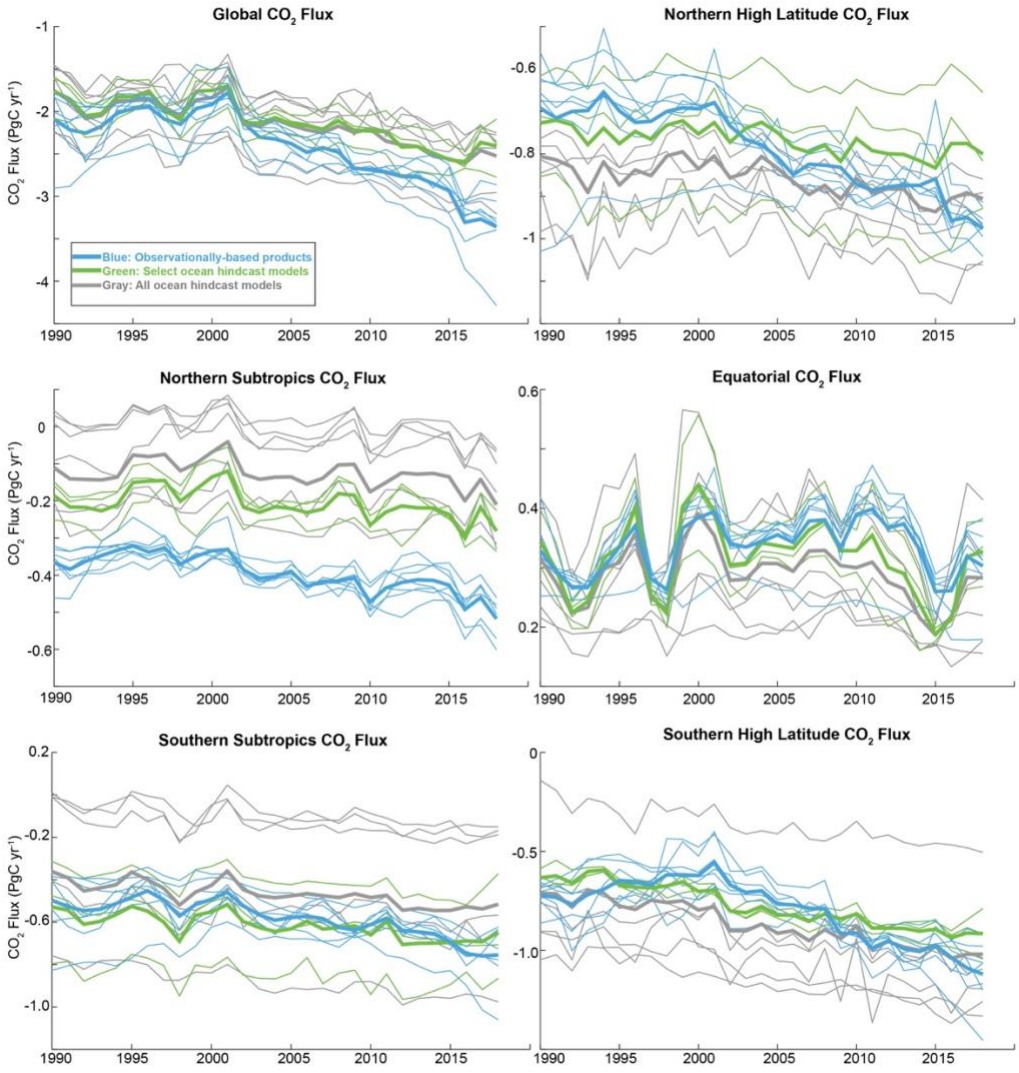

Figure S1. Time series for global and regionally-integrated air-sea CO2 fluxes (PgC/yr) from an ensemble of observation-based products (blue) and hindcast ocean models (gray and green), years 1990-2018. Thin lines are individual ensemble members while

bold lines show mean for the products (blue), the full ensemble of models (gray) and the selected models (green).

Observation-based pCO_2 products

Name	*Reference*
CMEMS-FFNN	*Denvil-Sommer et al. 2019 Chau et al. 2020*
CSIR-ML6	*Gregor et al. 2019*
Jena-MLS	*Rödenbeck et al. 2013*
JMA-MLR	*Iida et al. 2020*
LDEO-HPD	*Gloege et al. 2021*
MPI-SOMFFN	*Landschützer et al. 2014 Landschützer et al. 2020a*
NIES-FNN	*Zeng et al. 2014*

Hindcast Ocean Models

Name	*Reference*	Selected
CESM-ETHZ	*Doney et al. 2009*	Yes
CNRM-ESM2-1	*Berthet et al. 2019*	No
CSIRO	*Law et al. 2017*	No
FESOM-1.4-REcoM2	*Hauck et al. 2020*	No
MOM6-COBALT (Princeton)	*Liao et al. 2020*	No
MPIOM-HAMOCC6	*Paulsen et al. 2017*	No
NEMO-PISCES (IPSL)	*Aumont et al. 2015*	Yes
NEMO-PlankTOM5	*Buitenhuis et al. 2010*	Yes
NorESM-OCv1.2	*Schwinger et al. 2016*	No

Table S1. Observation-based pCO2 products and Hindcast Ocean Model names and references

Region	Percent of river efflux	River flux adjustment (% * 0.49PgC/yr)
Northern High Latitudes	10.3%	0.051 PgC yr^{-1}
Northern Subtropics	27.2%	0.133 PgC yr^{-1}
Equatorial	3.7%	0.018 PgC yr^{-1}
Southern Subtropics	35%	0.172 PgC yr^{-1}
Southern High Latitudes	6.2%	0.030 PgC yr^{-1}

Table S2. River adjustment from Lacroix et al. (2020)

Region	Lacroix % of river efflux	Aumont % of river efflux
90N-20N	34%	26%
20N-20S	45%	25%
20S-90S	21%	49%

Table S3. River adjustment from Lacroix et al. (2020) for latitude bounds as defined by Aumont et al. (2001).

	Products	Products natural outgassing due to rivers	3σ Product spread	Models	Select Models
Global	-1.94 ± 0.32	-2.43 ± 0.85	-3.36; -1.50	-2.15 ± 0.64	-2.11 ± 0.42
N High Latitudes	-0.74 ± 0.13	-0.79 ± 0.17	-1.01; -0.58	-0.86 ± 0.25	-0.77 ± 0.34
N Subtropics	-0.26 ± 0.04	-0.40 ± 0.18	-0.62; -0.17	-0.13 ± 0.22	-0.77 ± 0.09
Equatorial	0.35 ± 0.08	0.33 ± 0.10	0.21; 0.46	0.29 ± 0.14	0.32 ± 0.10
S Subtropics	-0.41 ± 0.16	-0.58 ± 0.34	-0.95; -0.21	-0.47 ± 0.60	-0.62 ± 0.45
S High Latitudes	-0.77 ± 0.10	-0.80 ± 0.13	-0.95; -0.65	-0.87 ± 0.48	-0.21 ± 0.07

Table S4 Global and regionally integrated air-sea CO_2 flux in PgC/yr for 7 observation-based products, products with river carbon efflux adjustment, all nine models, and select three models. Uncertainty (±) represents 2σ spread. Middle column shows the 3σ bounds on the product ensemble; this is used for the model selection criteria.

	Observation-based Products	Products with natural outgassing due to rivers	Hindcast Models	Select Models
Global	7%	13%	15%	10%
NH High Latitudes	8%	9%	15%	22%
NH Subtropics	5%	19%	88%	23%
Equatorial	12%	13%	24%	15%
SH Subtropics	14%	21%	65%	36%

SH High Latitudes	6%	6%	28%	5%

Table S5 Coefficient of Variability (CoV) for global and regionally integrated air-sea CO_2 fluxes.

	Models	Select Models
Northern High Latitudes	0.47	0.37
Northern Subtropics	0.12	0.09
Equatorial	0.04	0.04
Southern Subtropics	0.29	0.22
Southern High Latitudes	0.79	0.51

Table S6 Seasonal RMSE mean between ensemble of products and models or select models over five regions.

	Products: Models	Products: Select Models
Global	**0.71**	**0.64**
N High Latitudes	0.25	0.06
N Subtropics	**0.86**	**0.82**
Equatorial	**0.85**	**0.86**
S Subtropics	**0.50**	**0.53**
S High Latitudes	0.22	0.11

Table S7 Correlations for detrended CO_2 flux time series, between the product ensemble mean and that from all models and only select models. Values in bold are significant at the 95% confidence level (student t-test with $p<0.05$).