



RESEARCH ARTICLE

10.1029/2026MS005774

Evaluation of Timing and Amplitude Biases in the Controls of Simulated Southern Ocean $p\text{CO}_2$

Key Points:

- Most Earth System Models produce an out-of-phase seasonal CO_2 cycle south of the Subantarctic Front
- Compensating errors improve $p\text{CO}_2$ seasonal cycles in some models, giving the correct phasing but for the wrong mechanistic reasons
- Mixed layer depth biases are tightly linked to biases in dissolved inorganic carbon and temperature, and therefore a priority to address

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

S. M. Bushinsky,
seth.bushinsky@hawaii.edu

Citation:

Bushinsky, S. M., Arteaga, L. A., Fassbender, A. J., Hauck, J., Mazloff, M. R., Cervečki, I., et al. (2026). Evaluation of timing and amplitude biases in the controls of simulated Southern Ocean $p\text{CO}_2$. *Journal of Advances in Modeling Earth Systems*, 18, e2026MS005774. <https://doi.org/10.1029/2026MS005774>

Received 29 JAN 2026

Accepted 15 SEP 2026

Seth M. Bushinsky¹ , Lionel A. Arteaga^{2,3} , Andrea J. Fassbender⁴ , Judith Hauck^{5,6} , Matthew R. Mazloff⁷ , Ivana Cervečki⁷ , Peter Landschützer⁸ , Christopher Danek⁵ , Anastasia Romanou^{9,10} , Paul Lerner^{9,10} , Alison R. Gray¹¹ , and Sarah Schlunegger¹² 

¹Department of Oceanography, School of Ocean and Earth Science and Technology, University of Hawai'i at Mānoa, Honolulu, HI, USA, ²Global Modeling and Assimilation Office, NASA Goddard Space Flight Center, Greenbelt, MD, USA, ³Goddard Earth Sciences, Technology and Research II, University of Maryland Baltimore County, Baltimore, MD, USA, ⁴NOAA/OAR Pacific Marine Environmental Laboratory, Seattle, WA, USA, ⁵Alfred-Wegener-Institut, Helmholtz-Zentrum für Polar- und Meeresforschung, Bremerhaven, Germany, ⁶Department 02 Biology/Chemistry, University of Bremen, Bremen, Germany, ⁷Scripps Institution of Oceanography, University of California San Diego, La Jolla, CA, USA, ⁸Flanders Marine Institute (VLIZ), Ostend, Belgium, ⁹NASA Goddard Institute for Space Studies, New York, NY, USA, ¹⁰Department of Applied Physics and Applied Mathematics, Columbia University, New York, NY, USA, ¹¹School of Oceanography, University of Washington, Seattle, WA, USA, ¹²Program in Atmospheric and Oceanic Sciences, Princeton University, Princeton, NJ, USA

Abstract The Southern Ocean is a major sink for atmospheric carbon dioxide and critical to the current and future carbon cycle. This net annual CO_2 flux reflects the balance of strong seasonal variability characterized by opposing periods of winter outgassing and summer uptake. Using a simple framework, we evaluate how model biases in both the amplitude and timing of dissolved inorganic carbon (DIC) and total alkalinity (TA) and the amplitude of sea surface temperature (SST) impact simulated $p\text{CO}_2$. We examine seasonal CO_2 fluxes and $p\text{CO}_2$ south of the Subantarctic Front in 42 Earth System Model and three state estimate simulations. Only 11 of the 45 simulations have a seasonal $p\text{CO}_2$ cycle correlation of ≥ 0.7 to observed $p\text{CO}_2$, while 25 have a correlation of < 0 . Four of the well-correlated models accurately represent the seasonality of SST, DIC, and TA, while TA biases compensate for DIC or SST biases in the other seven. DIC and SST amplitude biases are related to mixed layer (MLD) biases, with shallow MLDs, especially in the summer, correlated with larger amplitude DIC and SST cycles than observed. The amplitude of seasonal Net Primary Production is correlated to DIC and TA timing. We assess the main adjustments to the phase and amplitude of control variables needed to correct the simulated $p\text{CO}_2$ seasonality in the evaluated models. These findings highlight the difficulty and importance of capturing the seasonal processes influencing the carbonate system to correctly model and predict the Southern Ocean carbon sink and its response to a changing climate.

Plain Language Summary The Southern Ocean takes up a large amount of the excess carbon released by human activities into the atmosphere. Computational models of the ocean are one of the main ways that we predict how the Southern Ocean carbon sink might change in the future. The seasonal exchange of carbon between the ocean and atmosphere is governed by opposing biological and physical cycles. These cycles are both very large and often work to counteract each other, so it can be difficult for models to capture the correct balance of physical and biological processes and therefore the air-sea carbon dioxide exchange. We focus on a large region of the Southern Ocean that is important to the uptake of carbon, finding that most of the models used to project the future of the Earth's carbon cycle do not correctly capture present seasonality. This means that they have the ocean taking up carbon in the winter when observations show it is releasing carbon and then do the opposite in the summer. We present a simple way to understand which ocean properties and processes are driving these discrepancies to guide future improvements to the representation of seasonal cycles in numerical models.

1. Introduction

The Southern Ocean (south of 35°S) accounts for approximately 40% of the contemporary carbon dioxide (CO_2) flux into the ocean each year (DeVries, 2014; Friedlingstein et al., 2023). This strong annual sink is the product of seasonally varying uptake and outgassing, driven by cycles of biology, mixing, and solubility (Rodgers et al., 2023). The air-sea CO_2 flux depends directly on the partial pressure of carbon dioxide ($p\text{CO}_2$) of seawater,

which is a function of dissolved inorganic carbon (DIC), sea surface temperature (SST), sea surface salinity (SSS), and total alkalinity (TA), with the highest influence from SST and DIC (Lauderdale et al., 2016; Takahashi et al., 2002). During the spring and summer growth period, surface biological production fixes inorganic carbon into organic carbon, lowering $p\text{CO}_2$, which will tend to draw carbon dioxide from the atmosphere into the ocean. In the wintertime, entrainment of deeper waters enriched in DIC from the accumulated respiration of sinking particles brings high $p\text{CO}_2$ water into contact with the atmosphere, which favors the outgassing of CO_2 . Opposing these biological and physical processes is the solubility cycle, with warming in the spring and summer decreasing the solubility of CO_2 , which increases the $p\text{CO}_2$ of surface waters. In the winter, heat loss to the atmosphere cools the surface waters, increasing the solubility and lowering $p\text{CO}_2$, favoring the uptake of CO_2 from the atmosphere (Gruber et al., 2019; Mikaloff Fletcher et al., 2007; Volk & Hoffert, 1985).

The balance of these competing processes determines the seasonal cycle of surface $\Delta p\text{CO}_2$ ($\Delta p\text{CO}_2 = p\text{CO}_{2,\text{seawater}} - p\text{CO}_{2,\text{atmosphere}}$) and therefore the direction of air-sea CO_2 fluxes and, in large part, the magnitude of the annual oceanic uptake of CO_2 . Future changes in ocean uptake of CO_2 will be due in part to rising atmospheric CO_2 driving increased oceanic uptake (or reduced oceanic outgassing) and in part to changes in the mechanisms that drive the seasonal cycle of surface $p\text{CO}_2$ (Fassbender et al., 2018). Earth System Models (ESMs) represent our best understanding of the processes controlling carbonate system dynamics and are used to evaluate how different emissions scenarios will impact global ocean carbon uptake and its associated feedbacks. Model-based projections suggest that the Southern Ocean will continue to play a major role in modulating global marine biogeochemical fluxes (Hauck et al., 2015; Kessler & Tjiputra, 2016; Terhaar et al., 2021). However, the accurate estimation of air-sea carbon fluxes in this region using models is hampered by uncertainties in the representation of its seasonal cycle (Anav et al., 2013; Mongwe et al., 2018, 2024). Observational constraints on the magnitude and trend of the CO_2 flux are also lacking, in large part due to the lack of wintertime observations (Friedlingstein et al., 2023; Gloege et al., 2021; Hauck, Gregor, et al., 2023; Rustogi et al., 2023).

Considerable effort has been devoted to assessing biases in model mean states (e.g., Terhaar et al., 2024). However, there is growing evidence that $p\text{CO}_2$ variations occurring on seasonal (Rustogi et al., 2023) and sub-seasonal (Carranza et al., 2024; Nickford et al., 2022) timescales have an important impact on the net annual air-sea CO_2 flux and that $p\text{CO}_2$ seasonal cycle amplitudes have changed significantly (Hauck & Völker, 2015; Hauck et al., 2015; Landschützer et al., 2018; Rodgers et al., 2008) and asymmetrically (Fassbender et al., 2018, 2022) over the past several decades. The importance of seasonality to the air-sea CO_2 flux balance means we need to understand sub-annual $p\text{CO}_2$ variability, especially since seasonality is not represented consistently across models (i.e., Rodgers et al., 2023). An inaccurate representation of the seasonal cycle indicates that a given model is failing to capture the correct interplay of physical, chemical, and biological cycles that govern contemporary CO_2 fluxes and thus it is unclear if that model can adequately project how changes in physical and biological drivers will impact future CO_2 fluxes (e.g., Gallego et al., 2018; Goris et al., 2018; Kessler & Tjiputra, 2016; Nevison et al., 2016). Given that the magnitude of the air-sea CO_2 flux is typically the residual of periods of outgassing and uptake, representation of the seasonal cycle may be an equally important metric for model skill in representing mechanisms as the magnitude of the annual flux.

Our primary observations of the seasonal cycle of surface $p\text{CO}_2$ come from underway shipboard measurements and mooring observations. These are rigorously quality controlled and compiled in the Surface Ocean CO_2 Atlas (SOCAT, Bakker et al., 2016, 2023). While extensive, these observations have gaps in space and time that are then interpolated using a range of mapping methods, including ones that relate the $p\text{CO}_2$ measurements to more widely available data, such as temperature, salinity, mixed layer depth (MLD), and satellite-derived chlorophyll concentration to create gridded maps. The gridded $p\text{CO}_2$ maps can then be used in air-sea carbon flux calculations (Fay et al., 2021; Rödenbeck et al., 2015). The density of the underlying shipboard observations is overall biased toward the Northern Hemisphere and summer months, leaving regions such as the Southern Ocean with only a few very localized wintertime observations to constrain the seasonal and annual fluxes. For example, from 2014 to 2018 there were 2.4x more $1^\circ \times 1^\circ$ ocean cells covered annually by SOCAT north of 30°N than south of 30°S , and 3x the coverage when considering winter months for each region (Hauck, Nissen, et al., 2023), this despite the southern region having over 2x more ocean cells.

Autonomous profiling floats have recently begun to fill in these wintertime data gaps. Floats equipped with pH sensors are being used to estimate $p\text{CO}_2$ year-round, especially in the Southern Ocean where over 300 floats have been deployed by the Southern Ocean Climate and Carbon Observations and Modeling project (SOCCOM,

Johnson et al., 2017; Sarmiento et al., 2023) and have further been implemented in gridded mapping (Bushinsky et al., 2019). Previous studies have shown that float-estimated $p\text{CO}_2$ is likely biased high, with direct comparisons between ship and float $p\text{CO}_2$ yielding biases of 2–5 μatm (Wu & Qi, 2023). Some of this bias can be traced to propagation of errors in float oxygen (Bushinsky et al., 2025), but the biases do not seem to impact the magnitude or timing of the seasonal cycle and therefore do not have an impact on the overall message of the work presented here. Between the SOCAT and SOCCOM data sets we are now able to more closely evaluate model representations of the Southern Ocean seasonal cycle. Rodgers et al. (2023) synthesized the representation of seasonal $p\text{CO}_2$ cycles in global ocean biogeochemistry models and ESMs, finding significant disagreements in amplitude and phase that were primarily due to DIC discrepancies. Here we expand on their analysis, including a larger range of ESMs and specifically focusing on how biases in phase (timing of the seasonal cycle) compound with amplitude biases to result in out of phase $p\text{CO}_2$ in many models.

Given the importance of accurate seasonal cycle representation in models, we provide a simple framework to diagnose the mechanisms driving $p\text{CO}_2$ seasonality biases in models. Our approach examines how seasonal phasing biases in DIC and TA combine with seasonal amplitude biases in DIC, TA, and SST to impact the phase of the $p\text{CO}_2$ cycle. We limit our analysis to the region between the Subantarctic Front to the north and 62°S to the south, a region where the presence of seasonal outgassing is key to the overall Southern Ocean carbon flux and therefore an important mechanistic test of Southern Ocean carbonate system dynamics (Bushinsky et al., 2019; Gray et al., 2018). The seasonal sea ice zone is an important contributor to the Southern Ocean carbon cycle, but we found that modeled air-sea carbon fluxes were often strongly influenced by the presence of extremely large polynyas that represent a fundamentally different set of processes than those we are considering here, leading to our 62°S boundary.

We analyze the seasonal cycle of 45 model simulations representing a range of Earth System Models and state estimates. In Section 3.1 we describe how well the models match observed seasonal cycles of CO_2 fluxes, $p\text{CO}_2$, DIC, SST, SSS, TA, MLD, and depth integrated net primary production (NPP_{int}). In Section 3.2 we use a harmonic fit of the seasonal carbonate system to iteratively perturb the system using the observed range of model biases to demonstrate how easily the seasonal cycle of $p\text{CO}_2$ can be pushed out of phase, thereby determining the approximate biases that can be tolerated in models while still reproducing a seasonal cycle that closely matches the timing of the observed $p\text{CO}_2$ cycle. We then use these results in Sections 3.3 and 3.4 to determine the corrections needed to bring model $p\text{CO}_2$ back into phase with observed $p\text{CO}_2$ and to explore how model parameters important to the seasonal cycle of the carbonate system vary relative to observations. Finally, we conclude with suggestions on how to use this information to inform efforts to improve existing model process representation.

2. Methods

2.1. Observational Data Sets and Products

The region between the Subantarctic Front and 62°S captures the area of the upwelling limb of overturning circulation that exhibits the strongest wintertime outgassing, which we call the Polar-Frontal Antarctic Zone (PFAZ). The Subantarctic Front location is from Gray et al. (2018) and defined as the boundary where potential temperature (θ) at 400 m equals 5°C. All observational data used are from 2010 to 2019. This time period is chosen due to the high number of observations of the entire seasonal cycle, including when a large number of SOCCOM floats are available and quality controlled.

Profiling floats equipped with pH and other biogeochemical sensors have been deployed in the Southern Ocean largely through the SOCCOM project. Float pH measurements can be used to estimate $p\text{CO}_2$ when paired with an estimate of TA (Bushinsky et al., 2019; Gray et al., 2018; Johnson et al., 2017; Maurer et al., 2021; Takeshita et al., 2018; Williams et al., 2017). These $p\text{CO}_2$ estimates have greater uncertainty than the shipboard measurements but complement the existing Surface Ocean CO_2 Atlas (SOCAT, Bakker et al., 2016) database by providing additional information about the seasonal cycle, especially during the winter and in open ocean regions away from common cruise tracks. Surface air-sea CO_2 fluxes and $p\text{CO}_2$ are from SOCCOM and SOCAT mapped products described in Bushinsky et al. (2019) using the Landschützer et al. (2013) and Rödenbeck et al. (2013) mapping methods, updated through 2022 using SOCATv2023 (Bakker et al., 2016, 2023) and the April 2023 SOCCOM data snapshot (Figure 1; Riser et al., 2023). Averages of the Landschützer et al. (2013) and Rödenbeck et al. (2013) SOCAT and SOCCOM + SOCAT products from 2010 to 2019 years are used in this study, labeled “Combined SOCAT” and “Combined SOCCOM + SOCAT.”

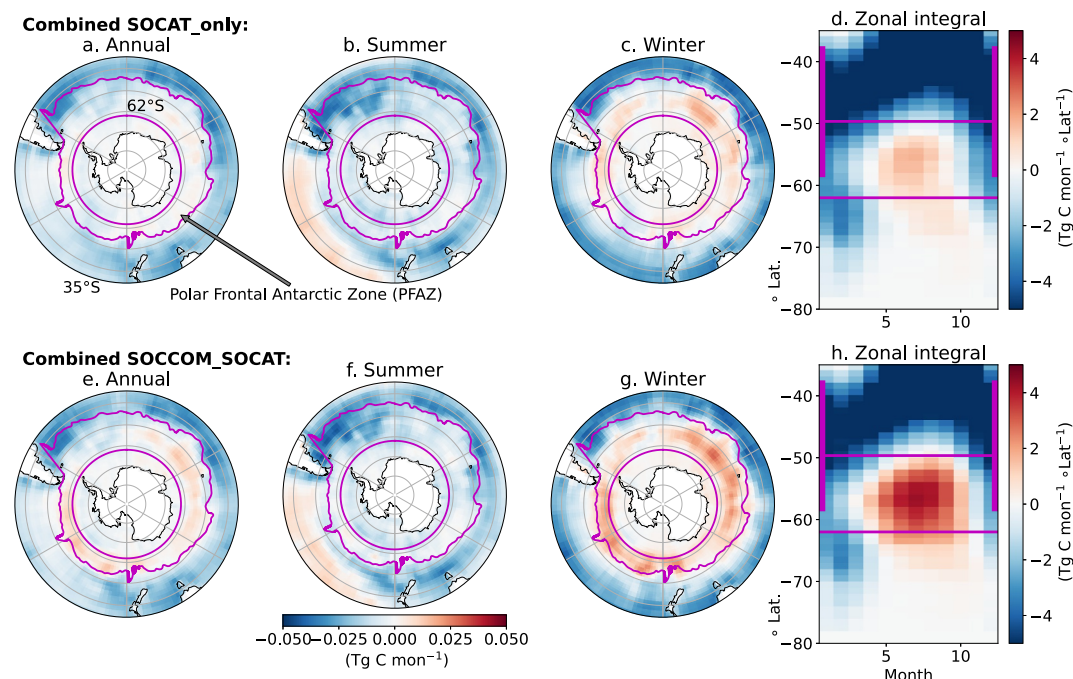


Figure 1. Air-sea carbon fluxes from observation-based products (positive to atmosphere). Top row, Combined SOCAT only (a) annual, (b) summer, (c) winter, and (d) zonal integrals, bottom row (e–f), the same but for the Combined SOCCOM + SOCAT product. “Combined” data shown are the averages of the SOM-FFN and Jena CarboScope products updated through 2022 (Landschützer et al., 2013; Rödenbeck et al., 2013). The northern magenta line is the Subantarctic Front (SAF, defined as the 5°C potential temperature (θ) contour at 400 m depth and calculated from core Argo observations), and the southern magenta line is 62°S. In (d and h) the mean latitude of the SAF is shown (northern magenta line), with the vertical lines on either end indicating the maximum/minimum latitudinal extent. Monthly zonal integrals indicate a seasonal cycle of summertime uptake transitioning to wintertime outgassing in the region between ~50°S to ~62°S. Wintertime outgassing is present in both data products, but strengthened when float $p\text{CO}_2$ estimates are included.

SSTs for 2010–2019 are from the monthly NOAA Optimal Interpolation SST v2 (Reynolds et al., 2002). SSS data are taken from the World Ocean Atlas 2023 database (Reagan et al., 2024). Mixed layer depths are calculated using a 0.03 kg m^{-3} potential density threshold from 10 m (de Boyer Montégut et al., 2004), based on core Argo temperature and salinity profiles (Roemmich & Gilson, 2009) that are objectively interpolated to a monthly $1^\circ \times 1^\circ$ grid. Using monthly gridded temperature and salinity yields a mean MLD that is shallower than the instantaneous mixing depth calculated from individual profiles but allows the same calculation from model output (de Boyer Montégut et al., 2004), removing possible discrepancies due to differences in algorithm or sampling.

DIC and TA data are from GLODAPv2.2023 (Key et al., 2015; Olsen et al., 2016), supplemented by DIC estimates derived from SOCCOM float pH (Gray et al., 2018; Williams et al., 2017, 2018). NPP_{int} is an average of three 2010–2019 NPP estimates downloaded from the “Ocean Productivity” site (<https://orca.science.oregonstate.edu/index.php>). The NPP estimates are derived from the Carbon-based-Productivity-Model (CbPM), Carbon, Absorption, and Fluorescence Euphotic-resolving (CAFE), and Vertically Generalized Production Model (VGPM) models applied to satellite observations (Behrenfeld & Falkowski, 1997; Behrenfeld et al., 2005; Silsbe et al., 2016; Westberry et al., 2008).

Monthly means from gridded data sets (CO_2 fluxes, $p\text{CO}_2$, SST, SSS, MLD, NPP_{int}) are calculated by selecting data within the PFAZ, masking out regions shallower than 1,000 m depth to avoid any discrepancies in how close each product goes to coastal regions, and then area weight averaging. Temporally-resolved data are detrended prior to calculating a mean annual cycle. In situ DIC and TA observations are separately binned to a $1^\circ \times 1^\circ$ grid for Argo and GLODAP, then averaged together to create a merged $1^\circ \times 1^\circ$ mean, and subsequently mean seasonal cycles over the region.

Table 1
Model Output Used in This Study

CMIP5	Variant	Ref.	CMIP6	Variant	Ref.
CESM1-BGC	r1i1p1	Long et al. (2013) and Moore et al. (2004)	ACCESS-ESM1-5 (6)	r1i1p1f1	Mackallah et al. (2022)
CMCC-CESM	r1i1p1	Vichi et al. (2007)	AWI-ESM-1-REcoM (6)	r1i1p1f1	Semmler et al. (2020), Danek and Hauck (2024), and Danek et al. (2023)
CNRM-CM5	r1i1p1	Séférián et al. (2013)	CESM2-WACCM (6)	r1i1p1f1	Meehl et al. (2020)
CanESM2	r1i1p1	Zahariev et al. (2008)	CESM2 (6)	r4i1p1f1	Meehl et al. (2020)
GFDL-ESM2G	r1i1p1	Dunne et al. (2012, 2013)	CMCC-ESM2 (6)	r1i1p1f1	Lovato et al. (2022)
GFDL-ESM2M	r1i1p1	Dunne et al. (2012, 2013)	CNRM-ESM2-1 (6)	r1i1p1f2	Séférián et al. (2019)
GISS-E2-H-CC	r1i1p1	Romanou et al. (2013) and Schmidt et al. (2014)	CanESM5-1 (6)	r1i1p1f1	Sigmond et al. (2023) and Christian et al. (2022)
GISS-E2-R-CC	r1i1p1	Romanou et al. (2013) and Schmidt et al. (2014)	CanESM5-CanOE (6)	r1i1p2f1	Sigmond et al. (2023) and Christian et al. (2022)
HadGEM2-CC	r2i1p1	Martin et al. (2011) and Palmer and Totterdell (2001)	CanESM5 (6)	r1i1p2f1	Swart et al. (2019)
HadGEM2-ES	r2i1p1	Martin et al. (2011) and Palmer and Totterdell (2001)	EC-Earth3-CC (6)	r1i1p1f1	Döscher et al. (2022)
IPSL-CM5A-LR	r1i1p1	Dufresne et al. (2013) and Marti et al. (2010)	GFDL-CM4 (6)	r1i1p1f1	Held et al. (2019)
IPSL-CM5A-MR	r1i1p1	Dufresne et al. (2013) and Marti et al. (2010)	GFDL-ESM4 (6)	r1i1p1f1	Dunne et al. (2020)
IPSL-CM5B-LR	r1i1p1	Dufresne et al. (2013) and Marti et al. (2010)	GISS-E2-1-G (6)	r10i1p1f1	Kelley et al. (2020) and Lerner et al. (2021)
MIROC-ESM-CHEM	r1i1p1	Watanabe et al. (2011)	INM-CM4-8 (6)	r1i1p1f1	Volodin et al. (2018)
MIROC-ESM	r1i1p1	Watanabe et al. (2011)	INM-CM5-0 (6)	r1i1p1f1	Volodin et al. (2018)
MPI-ESM-LR	r1i1p1	Giorgetta et al. (2013)	IPSL-CM6A-LR (6)	r1i1p1f1	Boucher et al. (2020)
MPI-ESM-MR	r1i1p1	Ilyina et al. (2013)	MIROC-ES2L (6)	r1i1p1f2	Hajima et al. (2020)
MRI-ESM1	r1i1p1	Adachi et al. (2013)	MPI-ESM1-2-HR (6)	r1i1p1f1	Mauritsen et al. (2019)
NorESM1-ME	r1i1p1	Tjiputra et al. (2013)	MPI-ESM1-2-LR (6)	r1i1p1f1	Mauritsen et al. (2019)
			MRI-ESM2-0 (6)	r1i2p1f1	Yukimoto et al. (2019)
			NorESM2-LM (6)	r1i1p1f1	Seland et al. (2020)
			NorESM2-MM (6)	r1i1p1f1	Seland et al. (2020)
			UKESM1-0-LL (6)	r1i1p1f2	Sellar et al. (2019)
State Estimates					
BSOSE-i133	–	Verdy and Mazloff (2017)			
BSOSE-i154	–	Verdy and Mazloff (2017)			
ECCO-v4r5-Darwin	–	Forget et al. (2015) and Carroll et al. (2020)			

2.2. Model Output

Forty two models from the Climate Model Intercomparison Projects (CMIP) 5 ($n = 19$) and 6 ($n = 23$) were evaluated for this study, along with two iterations of the Biogeochemical Southern Ocean State Estimate (BSOSE, Verdy & Mazloff, 2017) and one realization of the Estimating the Circulation and Climate of the Ocean version 4 release 5 coupled to the Darwin biogeochemical module (ECCOv4r5-Darwin; Carroll et al., 2020; Forget et al., 2015; Table 1). BSOSE and ECCOv4r5-Darwin are ocean-sea-ice state estimates forced with atmospheric reanalysis that assimilate observations but maintain closed budgets. Both BSOSE and ECCO assimilate physical data, while BSOSE also assimilates biogeochemical data. BSOSE iteration i133 covers only the years 2013–2018,

while i154 covers 2013–2023, and ECCOv4r5-Darwin includes the full 2010–2019 range. BSOSE iterations reflect on-going development and addition of new data assimilation, such that i154 is a newer version of the state estimate than i133. Experiment RCP 8.5 (years 2010–2019) was used for CMIP5 models and a concatenation of the historical (years 2010–2014) and SSP5-8.5 (2015–2019) experiments for CMIP6. RCP 8.5 and SSP5-8.5 are high emissions scenarios (Canadell et al., 2021) that most modeling centers have run and therefore biogeochemical output is available from a large number of models. Because the goal is to primarily compare present-day CO₂ fluxes to their model representation, atmospheric CO₂ in both RCP 8.5 and SSP5-8.5 can act a proxy for the modern climate even if the future emissions in these scenarios are high (Fuss et al., 2014; Peters et al., 2013). Model names are appended with “(6)” if from CMIP6. A single ensemble member was used for each model. While a single decade can have variability in individual model runs, we are looking at first order seasonal cycles, rather than other modes of variability.

CMIP models were chosen for this analysis if at least monthly CO₂ fluxes and monthly depth-resolved θ were available. Model data are pre-processed using Climate Data Operator (cdo; Schulzweida, 2023) to select target years, interpolate depths if necessary, set uniform calendars, reference times, and bilinearly remap output to 1° × 1° grids. All other processing is performed in Matlab and Python. Mean seasonality of each model variable was calculated from de-trended mean monthly seasonal cycles over the study period.

In contrast to some recent prior work (e.g., Mongwe et al., 2018) we use a region that spans across the Polar Front, rather than using the Polar Front as a boundary. We do this because the maximum wintertime outgassing in observational products and models is often centered at the Polar Front, but can also be north or south of it. Splitting the region at the Polar Front would then risk missing this flux in one region or the other, which might reduce the tight coupling we show later between DIC, SST, and MLDs. Additionally, for this reason we calculate the SAF separately for each model using the same criteria that is applied to observations in this study. In many of the flux maps for individual models (Figures S1–S3 in Supporting Information S1) the model-calculated SAF varies in location but closely follows the northern boundary of the mean annual outgassing region, highlighting the importance of using model-derived fronts instead of real-world boundaries to interpret model results.

Model MLDs are calculated from monthly gridded depth-resolved salinity and potential temperature using the same criteria as for observations. This provides a uniform MLD definition across all models and between models and observations.

To compare the wide range of modeled CO₂ fluxes to observed CO₂ flux seasonality, we average fluxes in the PFAZ for each model and calculate mean seasonal anomalies of CO₂ flux, $p\text{CO}_2$, DIC, SST, TA, SSS, NPP_{int}, and MLD for both observations/observation-based products and models (CO₂ flux, $p\text{CO}_2$, DIC, SST, and TA shown in Figure 2, first column; MLD, NPP_{int}, and SSS in Figure S7 in Supporting Information S1). We plot the relationship between the observed and modeled seasonal cycles in normalized Taylor diagrams (Figure 2 and Figure S7 in Supporting Information S1, second and third columns; fit values for all models and variables in Table 2). In a normalized Taylor diagram, the radial distance from the origin gives the relative standard deviation (SD) between a model and reference, which in this case is observations or an observation-based product (σ_m/σ_o) (Taylor, 2001). For a sine curve, this relative SD is equivalent to the amplitude ratio between the model and observed. The angle between the x -axis and each point is the phase shift (α), calculated from the correlation (R) between the model field and observed:

$$\alpha = \cos^{-1} R \quad (1)$$

Finally, the root mean square difference (E) can be calculated from the difference between the reference point (relative amplitude of 1, angle of 0) and a given model point:

$$E = \sqrt{[(\sigma_m/\sigma_o)\cos(\alpha) - 1]^2 + [(\sigma_m/\sigma_o)\sin(\alpha)]^2} \quad (2)$$

This gives a straightforward way to compare the seasonal progression of CO₂ fluxes, $p\text{CO}_2$, and the parameters that control $p\text{CO}_2$. For most of our analysis we focus on seasonal anomalies, rather than the absolute value, though mean state biases are also present in models and can impact $p\text{CO}_2$ (Fu et al., 2022; Terhaar et al., 2024). Interannual variability in each model's seasonal cycles over the 10 years analyzed is shown by first calculating the

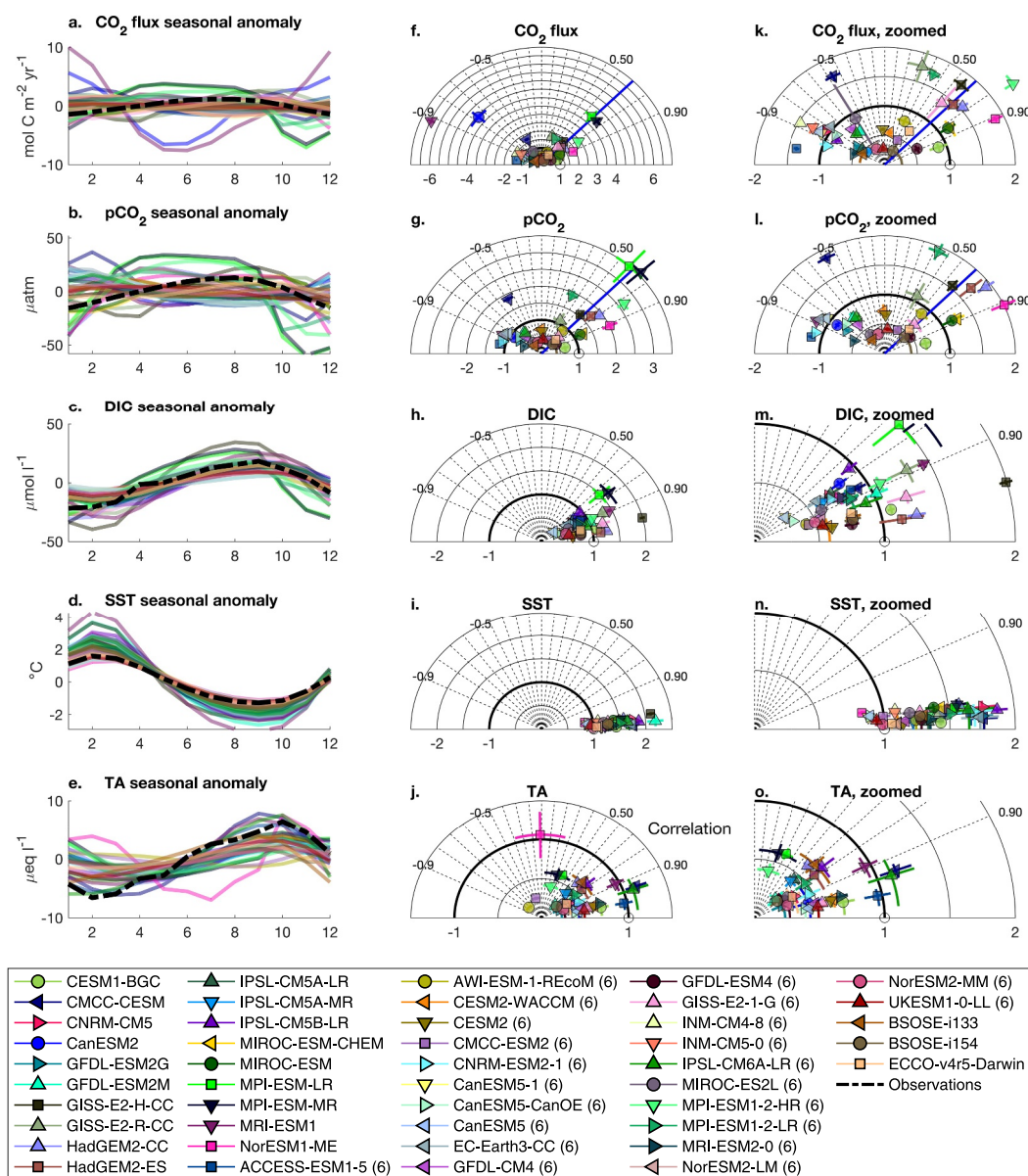


Figure 2. Seasonal anomalies of spatially averaged air-sea CO_2 fluxes, $p\text{CO}_2$, dissolved inorganic carbon (DIC), sea surface temperature (SST), and TA between the SAF and 62°S (PFAZ) in models and observations/observation-based products. First column: Seasonal anomalies of observations (black dashed lines) and model output (colored lines). Second column: Model fit to observations is plotted in Taylor diagrams with normalized standard deviation, σ_m/σ_o (SD), plotted as distance from the origin and correlation represented by the angle from the x-axis. Standard deviations of normalized SD and correlation from the 10 model years used are plotted as error bars around each marker. Observed seasonal cycles are taken from mapped CO_2 products (CO_2 flux and $p\text{CO}_2$), ship and float observations (DIC, TA), and satellite-derived measurements (SST). Observations are represented by the open black circle (normalized SD of 1, correlation of 1). Note the different scale for normalized SD in each Taylor diagram. Third column: the same Taylor diagrams as column 2, but zoomed in on data points close to observations. In (f, g, k, and l) the solid blue line indicates a correlation of 0.7.

correlation and normalized SD for each year, then plotting the standard deviations as error bars around each point in the second and third columns of Figure 2. Note that this is not meant to be representative of the overall interannual variability of each model, but just to show the degree to which these metrics vary during our study period.

Table 2

Model Mean Seasonal Cycle Output Correlation (R), Normalized Standard Deviation (σ_m/σ_o), and Normalized Root Mean Square Error (E) Relative to Observations/
Observation-Based Product Seasonal Cycles^a

	CO ₂ flux			$p\text{CO}_2$			DIC			SST			SSS			TA			MLD			NPP		
	R	$\frac{\sigma_m}{\sigma_o}$	E	R	$\frac{\sigma_m}{\sigma_o}$	E	R	$\frac{\sigma_m}{\sigma_o}$	E	R	$\frac{\sigma_m}{\sigma_o}$	E	R	$\frac{\sigma_m}{\sigma_o}$	E	R	$\frac{\sigma_m}{\sigma_o}$	E	R	$\frac{\sigma_m}{\sigma_o}$	E	R	$\frac{\sigma_m}{\sigma_o}$	E
CMIP5																								
CESM1-BGC	0.9	1.0	0.3	1.0	0.7	0.4	1.0	1.1	0.3	1.0	1.4	0.4	1.0	1.9	0.9	1.0	0.7	0.3	1.0	0.7	0.3	0.9	2.1	1.2
CMCC-CESM	−0.5	1.8	2.4	−0.5	1.9	2.5	0.9	0.9	0.5	1.0	1.8	0.8	1.0	2.5	1.5	0.9	1.2	0.4	1.0	0.5	0.5	0.6	1.6	1.3
CNRM-CM5	−0.8	0.8	1.7	—	—	—	0.8	0.7	0.6	1.0	1.8	0.8	1.0	1.2	0.3	0.9	0.4	0.7	1.0	0.6	0.4	0.7	1.1	0.8
CanESM2	−0.8	4.7	5.5	−0.8	0.9	1.8	0.8	0.8	0.6	1.0	1.5	0.5	1.0	1.3	0.3	1.0	0.4	0.6	1.0	0.6	0.4	0.9	0.7	0.5
GFDL-ESM2G	−0.8	0.5	1.4	−0.8	0.7	1.6	0.9	0.8	0.4	1.0	1.7	0.8	1.0	1.0	0.2	0.9	0.2	0.8	1.0	0.8	0.3	0.9	2.4	1.5
GFDL-ESM2M	−0.6	0.6	1.4	−0.6	0.8	1.6	0.9	1.0	0.4	1.0	2.2	1.2	1.0	1.3	0.4	0.7	0.2	0.9	1.0	0.7	0.3	0.9	2.4	1.6
GISS-E2-H-CC	0.7	1.2	0.9	0.7	1.6	1.2	1.0	2.0	1.1	1.0	2.1	1.1	0.9	3.4	2.5	—	—	—	0.9	1.0	0.3	0.9	1.1	0.5
GISS-E2-R-CC	0.3	1.1	1.3	0.5	1.1	1.1	0.9	1.3	0.6	1.0	1.4	0.4	0.9	1.1	0.4	—	—	—	1.0	1.6	0.7	0.9	1.5	0.7
HadGEM2-CC	0.8	1.2	0.7	0.8	1.9	1.3	1.0	1.3	0.3	1.0	1.6	0.6	1.0	1.6	0.6	0.7	0.6	0.7	1.0	0.8	0.3	0.9	4.4	3.4
HadGEM2-ES	0.7	1.1	0.8	0.8	1.7	1.2	1.0	1.1	0.2	1.0	1.5	0.5	1.0	1.6	0.7	0.8	0.6	0.7	1.0	0.7	0.3	0.9	4.2	3.3
IPSL-CM5A-LR	−0.8	0.6	1.5	—	—	—	0.8	0.7	0.6	1.0	1.3	0.3	0.9	1.2	0.6	0.7	0.4	0.8	1.0	0.6	0.4	0.7	1.1	0.9
IPSL-CM5A-MR	−0.7	0.4	1.3	—	—	—	0.8	0.6	0.6	1.0	1.2	0.2	0.9	1.2	0.6	0.7	0.4	0.8	1.0	0.5	0.5	0.7	1.1	0.9
IPSL-CM5B-LR	−0.9	0.6	1.5	—	—	—	0.8	1.0	0.7	1.0	1.9	0.9	0.9	1.7	0.9	0.8	0.7	0.6	1.0	0.5	0.6	0.7	1.2	0.9
MIROC-ESM-CHEM	0.8	1.2	0.7	0.9	1.3	0.6	—	—	—	1.0	1.4	0.5	1.0	1.4	0.5	—	—	—	1.0	0.5	0.5	0.9	2.6	1.7
MIROC-ESM	0.8	1.2	0.6	0.9	1.2	0.6	—	—	—	1.0	1.4	0.4	0.9	1.3	0.5	—	—	—	1.0	0.5	0.5	0.9	2.7	1.7
MPI-ESM-LR	0.7	2.6	2.0	0.7	3.5	2.9	0.7	1.5	1.0	1.0	1.5	0.6	1.0	1.5	0.6	0.4	0.6	0.9	0.9	0.5	0.5	0.6	5.6	5.0
MPI-ESM-MR	0.8	3.1	2.4	0.7	3.6	2.9	0.8	1.6	1.1	1.0	1.6	0.6	1.0	1.5	0.6	0.3	0.6	1.0	0.9	0.5	0.5	0.7	5.7	5.1
MRI-ESM1	−0.9	6.3	7.2	−0.9	1.2	2.1	0.9	1.5	0.7	1.0	2.6	1.6	1.0	2.6	1.6	0.9	1.0	0.5	1.0	0.7	0.4	0.9	1.0	0.4
NorESM1-ME	0.9	2.0	1.1	0.9	2.0	1.2	0.9	0.7	0.5	1.0	0.8	0.2	1.0	0.8	0.2	−0.0	1.1	1.5	1.0	0.8	0.3	0.9	5.7	4.8
CMIP6																								
ACCESS-ESM1-5 (6)	−1.0	1.0	2.0	−1.0	1.2	2.1	0.9	0.9	0.5	1.0	1.7	0.7	1.0	2.1	1.1	1.0	1.0	0.2	1.0	0.6	0.4	0.9	0.9	0.4
AWI-ESM-1-REcoM (6)	0.4	0.9	1.0	0.6	0.9	0.8	0.9	0.4	0.6	1.0	1.2	0.3	1.0	1.5	0.5	−0.7	0.2	1.2	1.0	0.7	0.3	0.6	1.5	1.2
CESM2-WACCM (6)	0.1	0.6	1.1	−0.0	0.7	1.2	1.0	0.6	0.5	1.0	1.1	0.1	1.0	1.3	0.3	1.0	0.6	0.4	1.0	0.8	0.3	0.9	1.8	1.1
CESM2 (6)	−0.0	0.6	1.2	0.0	0.7	1.2	1.0	0.6	0.4	1.0	1.1	0.2	1.0	1.2	0.3	1.0	0.6	0.4	1.0	0.9	0.2	0.9	1.9	1.2
CMCC-ESM2 (6)	0.4	0.4	0.9	0.4	0.4	0.9	0.9	0.4	0.6	1.0	1.0	0.1	0.9	0.7	0.4	−0.3	0.3	1.1	1.0	0.7	0.3	0.6	1.2	1.0
CNRM-ESM2-1 (6)	−0.9	0.8	1.8	−0.9	1.1	2.1	0.9	0.7	0.5	1.0	1.7	0.7	1.0	1.0	0.2	0.9	0.5	0.6	1.0	0.7	0.4	0.6	0.7	0.8
CanESM5-1 (6)	−0.9	1.0	2.0	−0.8	1.1	2.0	0.8	0.3	0.8	1.0	0.9	0.2	1.0	1.0	0.1	1.0	0.5	0.6	1.0	0.7	0.3	0.9	0.3	0.7
CanESM5-CanOE (6)	−0.8	0.8	1.7	−0.8	0.8	1.7	0.9	0.3	0.7	1.0	0.9	0.2	1.0	0.9	0.1	1.0	0.4	0.6	1.0	0.7	0.3	0.9	0.7	0.4
CanESM5 (6)	−0.9	1.1	2.0	−0.8	1.1	2.0	0.7	0.3	0.8	1.0	0.9	0.2	1.0	0.9	0.1	1.0	0.4	0.6	1.0	0.7	0.3	0.9	0.3	0.7
EC-Earth3-CC (6)	−0.8	0.6	1.5	−0.9	1.1	2.1	0.8	0.6	0.6	1.0	1.7	0.7	0.9	1.0	0.4	—	—	—	1.0	0.5	0.5	0.5	0.9	0.9
GFDL-CM4 (6)	−0.6	0.8	1.7	−0.6	0.6	1.5	0.9	0.6	0.5	1.0	1.2	0.2	1.0	1.2	0.2	1.0	0.5	0.5	1.0	0.9	0.2	0.4	1.4	1.4
GFDL-ESM4 (6)	0.9	0.6	0.6	0.8	0.5	0.7	1.0	0.8	0.3	1.0	1.3	0.3	1.0	1.1	0.2	0.9	0.3	0.8	1.0	0.6	0.4	0.9	2.3	1.5
GISS-E2-1-G (6)	0.6	1.5	1.1	0.7	0.7	0.7	1.0	1.2	0.4	1.0	1.3	0.3	1.0	0.9	0.2	1.0	0.5	0.5	1.0	0.9	0.1	0.9	1.9	1.1
INM-CM4-8 (6)	−0.9	1.3	2.2	—	—	—	—	—	—	1.0	1.1	0.2	0.9	0.6	0.6	—	—	—	0.9	0.4	0.6	—	—	—
INM-CM5-0 (6)	−0.9	1.3	2.2	—	—	—	—	—	—	1.0	1.1	0.2	0.9	0.7	0.5	—	—	—	0.9	0.4	0.6	—	—	—
IPSL-CM6A-LR (6)	−0.6	0.6	1.4	−0.6	0.7	1.6	0.9	0.9	0.4	1.0	1.6	0.7	1.0	2.4	1.5	0.9	1.1	0.4	1.0	0.6	0.4	0.7	1.1	0.8
MIROC-ES2L (6)	−0.5	0.6	1.4	−0.6	0.4	1.3	0.9	0.6	0.6	1.0	1.2	0.2	0.9	0.7	0.5	0.8	0.3	0.8	1.0	0.4	0.6	0.9	1.6	0.8

Table 2
Continued

	CO ₂ flux			pCO ₂			DIC			SST			SSS			TA			MLD			NPP		
	<i>R</i>	$\frac{\sigma_m}{\sigma_o}$	<i>E</i>	<i>R</i>	$\frac{\sigma_m}{\sigma_o}$	<i>E</i>	<i>R</i>	$\frac{\sigma_m}{\sigma_o}$	<i>E</i>	<i>R</i>	$\frac{\sigma_m}{\sigma_o}$	<i>E</i>	<i>R</i>	$\frac{\sigma_m}{\sigma_o}$	<i>E</i>	<i>R</i>	$\frac{\sigma_m}{\sigma_o}$	<i>E</i>	<i>R</i>	$\frac{\sigma_m}{\sigma_o}$	<i>E</i>	<i>R</i>	$\frac{\sigma_m}{\sigma_o}$	<i>E</i>
MPI-ESM1-2-HR (6)	0.8	1.9	1.2	0.8	2.7	1.9	0.9	1.1	0.5	1.0	1.3	0.3	1.0	1.2	0.3	0.3	0.4	1.0	1.0	0.6	0.5	0.8	3.7	3.0
MPI-ESM1-2-LR (6)	0.4	1.0	1.1	0.4	1.9	1.7	0.9	0.8	0.5	1.0	1.5	0.5	1.0	1.4	0.4	0.8	0.5	0.7	0.9	0.5	0.5	0.6	3.0	2.5
MRI-ESM2-0 (6)	−0.9	0.5	1.5	−0.9	0.5	1.5	0.9	0.8	0.5	1.0	1.4	0.4	1.0	1.5	0.5	1.0	0.7	0.4	1.0	0.9	0.2	0.9	0.9	0.4
NorESM2-LM (6)	−0.4	0.3	1.2	−0.2	0.4	1.2	0.9	0.5	0.5	1.0	1.1	0.1	1.0	0.6	0.4	0.9	0.3	0.7	0.9	0.2	0.8	0.9	1.6	0.8
NorESM2-MM (6)	−0.4	0.2	1.1	−0.3	0.3	1.1	0.9	0.5	0.6	1.0	1.0	0.0	1.0	0.6	0.5	0.9	0.3	0.8	1.0	0.5	0.5	0.9	1.5	0.7
UKESM1-0-LL (6)	−0.1	0.2	1.0	0.1	0.4	1.0	1.0	0.5		1.0	0.9	0.1	1.0	0.8	0.2	1.0	0.5	0.5	1.0	0.7	0.4	0.9	2.1	1.4
State Estimates																								
BSOSE-i133	−0.9	0.4	1.3	−0.6	0.3	1.2	1.0	0.8	0.3	1.0	1.3	0.3	0.9	1.7	0.9	0.7	0.6	0.7	1.0	0.6	0.4	0.9	1.6	0.9
BSOSE-i154	0.5	0.3	0.9	0.7	0.4	0.8	1.0	0.8	0.3	1.0	1.3	0.3	1.0	1.0	0.2	0.8	0.2	0.8	1.0	0.8	0.3	0.9	1.3	0.7
ECCO-v4r5-Darwin	0.6	0.7	0.9	0.7	0.6	0.7	1.0	0.8	0.4	1.0	1.1	0.1	1.0	1.1	0.3	0.9	0.3	0.7	1.0	0.8	0.3	0.9	1.2	0.5

^aCells are shaded from orange (worse agreement to observations) to purple (better), normalized to each individual column. Normalization was done because the significance of a discrepancy in a given metric can have very different importance depending on the parameter. For instance, a 0.9 correlation for MLD is poor relative to most other models, while 0.9 correlation for pCO₂ would be outstanding. This normalization also means that in cases where all values are similar, for instance SST

2.3. Testing pCO₂ Sensitivity to Seasonal Cycle Biases With a Harmonic Fit

To explore the impact of changes in seasonal SST, DIC, and TA on the pCO₂ correlation between models and observational products, we create a harmonic fit based on the mean seasonal cycles of SST, DIC, TA, and SSS from observations and perturb SST, DIC, and TA in ways representative of model biases (see Text S1 in Supporting Information S1). It is common to categorize pCO₂ in regions of the real or modeled ocean as driven by DIC or SST by calculating the impact of DIC or SST on pCO₂ while holding the other constant at a mean annual value (Fassbender et al., 2018; Mongwe et al., 2018; Prend et al., 2022; Takahashi et al., 2009). However, while this previous work primarily focused on how DIC or SST amplitudes dominated the seasonal cycle of pCO₂, we expand on this approach by also exploring how shifts in DIC, SST, or TA seasonal timing interact with amplitude biases to impact the seasonal pCO₂ cycle and correlation to observed pCO₂.

We fit the first harmonic of observed mean seasonal cycles in pCO₂, TA, SST, and SSS:

$$f(t) = a \cos\left(2\pi \frac{t}{L} + \phi\right) + b \quad (3)$$

where the harmonic is the least squares fit of the function ($f(t)$) to the observed seasonal cycle, a is the amplitude, t is the time (day of year), L is the period in days (i.e., $365.25/n$, where n is the order of the harmonic), ϕ is the phase (seasonal timing expressed in radians), and b is the mean seasonal offset from 0. To create the harmonics we solve for a , ϕ , and b . All carbonate system calculations are done using the CO2SYS Matlab package (van Heuven et al., 2011) using the carbonate system constants from Williams et al. (2017) (K_1 and K_2 dissociation constants of Lueker et al. (2000), HSO_4^- constant of Dickson (1990), HF constant of Perez and Fraga (1987) and the Lee et al. (2010) boron to salinity ratio). We calculate a DIC seasonal cycle from the pCO₂ and TA harmonics. The median monthly anomaly difference between DIC determined from GLODAP and float observations and DIC calculated as described above is only $\sim 1 \mu\text{mol kg}^{-1}$, but we use DIC calculated from pCO₂ and TA instead of the observed DIC to ensure an internally consistent carbonate system for our harmonic fit. DIC observations are largely from GLODAP, while pCO₂ is from the combined mapped products, so small differences between the two could introduce biases that are not relevant to our question here. The resulting internally consistent harmonic seasonal cycles are then used in two ways. First, they are used to test the sensitivity of seasonal pCO₂ to biases in

DIC, SST, and TA (Section 3.2) and second, as a target for adjusting model mean seasonal cycles in the PFAZ and assessing the response of $p\text{CO}_2$ correlations, as described below.

After fitting seasonal harmonics to observational products of $p\text{CO}_2$, SST, SSS, and DIC, we can then calculate corresponding model harmonics, enabling direct comparison between the two. As with observations, we fit a harmonic to the seasonal cycles of each ESM or state estimate's SST, DIC, TA, SSS, and $p\text{CO}_2$. We sequentially adjust the first harmonic of the seasonal cycles of model DIC, SST, and/or TA by shifting the timing and/or increasing or decreasing the amplitude according to the calculated difference between the first harmonics of the modeled and observed seasonal cycles. With the original SSS and the adjusted DIC, SST, and/or TA we calculate a new $p\text{CO}_2$ seasonal cycle and re-determine the correlation, normalized SD (amplitude ratio), and normalized root mean square error between the adjusted model $p\text{CO}_2$ and harmonic of the observed $p\text{CO}_2$. This adjustment procedure is designed to identify which physical or chemical processes contribute most to seasonal biases in model $p\text{CO}_2$ and improve our understanding of the sources of model error. We chose to adjust these three parameters (DIC, SST, and TA) because these constrain the carbonate system, solubility, and therefore surface $p\text{CO}_2$. SSS is also important for solubility, but the sensitivity of $p\text{CO}_2$ to model SSS biases was small enough that we ignored it for the purposes of this analysis.

For an example of iteratively adjusting model biases, the CanESM2 $p\text{CO}_2$ harmonic starts with a $p\text{CO}_2$ correlation ($\text{Corr}_{p\text{CO}_2}$) to the harmonic calculated from the observation-based $p\text{CO}_2$ product of -0.8 (Figure 3d). The DIC cycle in CanESM2 is somewhat out of phase with observations, with a wintertime peak 28 days later than observed. Correcting this by shifting the DIC harmonic by -28 days (moved earlier) improves the $p\text{CO}_2$ timing, bringing the correlation to 0.6 (Figure 3h, purple line). If we instead adjust only for the DIC amplitude difference of $-3.3 \mu\text{mol l}^{-1}$, this brings the $\text{Corr}_{p\text{CO}_2}$ to -0.3 from -0.8 . By adjusting both the DIC timing and amplitude together, the $\text{Corr}_{p\text{CO}_2}$ improves to 1 from the original -0.8 . In Section 3.3, we demonstrate how this adjustment approach can be used in practice to untangle model biases.

3. Results and Discussion

3.1. Modeled Carbonate System Seasonal Cycles

As found in Bushinsky et al. (2019), wintertime CO_2 fluxes to the atmosphere south of the SAF are a central feature in both the Combined SOCAT and Combined SOCAT + SOCCOM products, with winter outgassing enhanced when float observations are added to the mapped products (Figure 1). While there is substantial zonal asymmetry in the air-sea CO_2 flux, a zonal integral provides a simple synopsis of the annual cycle, revealing a seasonal switch from summertime uptake to a bullseye of outgassing during the winter from $\sim 50^\circ\text{S}$ to $\sim 62^\circ\text{S}$. North of the SAF, or zonally between 35°S and 45°S , the ocean follows an opposite pattern, with near neutral or slight summertime outgassing giving way to strong winter uptake.

Annual mean CO_2 flux maps (Figures S1–S3 in Supporting Information S1) and monthly zonal integrals (Figures S4–S6 in Supporting Information S1) from the models analyzed for this study show a wide range of CO_2 fluxes within the Southern Ocean, with some models demonstrating an annual mean CO_2 flux pattern similar to observed mean seasonal outgassing between the SAF and 62°S . However, many models that have somewhat similar spatial patterns in the annual mean have a flipped seasonal cycle that is evident from the monthly zonal integrals (ex. GFDL-ESM2G/ESM2M), or much stronger seasonal magnitudes of uptake/outgassing (ex. MPI-ESM-LR/MR), or both a flipped seasonal cycle and stronger fluxes (ex. MRI-ESM1).

Eleven models have a CO_2 flux correlation of ≥ 0.7 , which we arbitrarily set as a target threshold for a “good” correlation for CO_2 (Table 2, Figures 2f, 2g, 2k, and 2l, blue line), while the other 34 had correlations < 0.7 . For reference, a correlation of 0.7 equates to approximately 1 month difference in timing of CO_2 flux or $p\text{CO}_2$. 9 of the 11 models with a CO_2 flux correlation ≥ 0.7 were from CMIP5 and 2 from CMIP6. 25 out of 45 models had a CO_2 flux correlation of < 0 , with an overall range of -0.98 to 0.94 (mean of -0.14).

The CO_2 flux is driven by the air-sea disequilibrium of $p\text{CO}_2$ and a wind-speed dependent gas transfer velocity (e.g., Wanninkhof, 2014). Differences in wind-speed magnitude or location can be a source of error in CO_2 fluxes, especially during localized storm events (Carranza et al., 2024). However, $p\text{CO}_2$ normalized RMSE, correlation, and normalized SD are all highly correlated to model CO_2 fluxes, with the linear regression between regionally averaged seasonal model CO_2 flux and $p\text{CO}_2$ having an R^2 of 0.98 (Figure S8 in Supporting Information S1). This

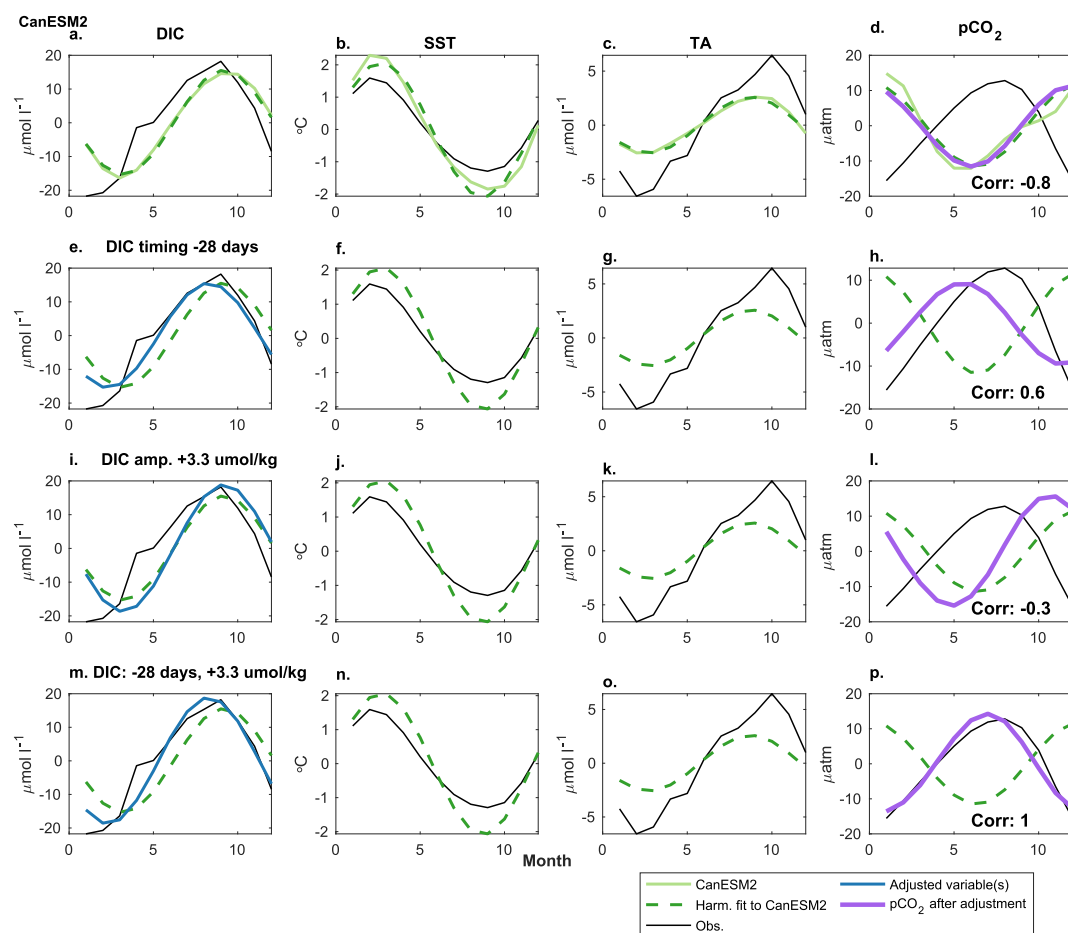


Figure 3. Example of adjusting CanESM2 dissolved inorganic carbon (DIC) to improve $p\text{CO}_2$ correlation. First, mean seasonal cycles (light green line (a–d)) of model DIC, sea surface temperature (SST), TA, SSS (SSS not shown) and $p\text{CO}_2$ in the PFAZ are fit with a harmonic (dashed green line in all panels). $p\text{CO}_2$ is calculated from model harmonic fits ((d), purple line) which closely matches the original model $p\text{CO}_2$. The correlation between calculated model $p\text{CO}_2$ and observed $p\text{CO}_2$ is originally -0.8 . DIC timing is first adjusted by the difference in phase between CanESM2 DIC and observed DIC (-28 days or the seasonal cycle shifted earlier, blue line in (e)) to match observed DIC timing. $p\text{CO}_2$ is then recalculated from the adjusted DIC and original model harmonic fits, which leads to a correlation improvement to 0.6 from -0.8 (h). In row 3, DIC amplitude is increased by $3.3 \mu\text{mol kg}^{-1}$ to match the observed DIC amplitude and in row 4 both DIC timing and amplitude are adjusted. TA and SST seasonal cycles are shown here to demonstrate the magnitude of variability for all important parameters, while full adjustments for DIC, SST, and TA for all models are shown in Figures S29–S59 in Supporting Information S1.

indicates that $p\text{CO}_2$ biases are the dominant issue for matching monthly seasonal CO_2 fluxes and we can ignore model wind speed biases for the purposes of this study.

The seasonal cycle of modeled DIC is in better agreement with observations than the CO_2 flux or $p\text{CO}_2$, with a range in correlation from 0.74 to 0.99 (mean of 0.89), but with a normalized SD range of 0.31 – 2 (mean of 0.85). The timing of SST is even more closely matched to observations (correlation from 0.99 to 1 , mean of 1), though with a normalized SD range of 0.84 – 2.6 , with most models displaying a larger seasonal amplitude than observed (mean of 1.4). SSS is less well correlated than SST (Figure S7 in Supporting Information S1), with a range in correlation from 0.85 to 1 (mean 0.96) and also has a wider normalized SD range (0.55 – 3.4 , mean of 1.3), but has less of an impact on CO_2 solubility and therefore on the seasonal cycle of $p\text{CO}_2$. Modeled TA has a very wide range of correlation to observations, from -0.74 to 0.98 (mean of 0.74), with normalized SDs that are on average approximately half of the observed (mean of 0.53 , range from 0.2 to 1.2). Most models have a MLD that is well correlated to that calculated from Argo float observations (correlation range of 0.92 – 1 , mean of 0.98), with a normalized SD that is usually smaller than observed (range of 0.44 – 1.6 , mean of 0.67). Finally, NPP_{int} has a correlation range of 0.42 – 0.94 (mean 0.81) with a normalized SD to observations of 0.31 – 5.7 and a mean of 2 .

Interestingly, splitting and comparing air-sea carbon fluxes based purely on latitudinal boundaries can lead to conflicting interpretations. For example, CESM1-BGC has been previously shown to match observed seasonal cycles of air-sea oxygen and carbon dioxide fluxes in 45–55°S based on atmospheric data (Nevison et al., 2016), which agrees with our findings that it is one of the best models in representing PFAZ CO₂ fluxes. However, Nevison et al. (2016) also found that GFDL-ESM2M, GFDL-ESM2G, and IPSL-CM5A-LR were top-performing models, while MPI-ESM-LR, NorESM1, and HadGEM2 were all out of phase with respect to CO₂ fluxes. This at first appears opposite to our analysis, with GFDL-ESM2M, GFDL-ESM2G, and IPSL-CM5A-LR all at a CO₂ flux correlation of -0.6 or worse and MPI-ESM-LR, NorESM1-ME, and both versions of HadGEM2 at 0.7 or better. However, looking at the Hovmoeller seasonal CO₂ fluxes (Figures S4 and S5 in Supporting Information S1) for those models makes it clear that the 45–55°S region spans both the northern extent of wintertime outgassing to the south and the southern extent of wintertime uptake to the north, explaining the contrary results. Our analysis based on the location of the individually modeled SAF effectively separates these two regions and yields more insight into the seasonal cycle of model processes.

3.2. Sensitivity of Seasonal $p\text{CO}_2$ Cycle to Biases in DIC, SST and TA

We target the $p\text{CO}_2$ correlation between models and observation-based products because of the wide variability in the timing of modeled $p\text{CO}_2$ and due to previous work demonstrating the importance of CO₂ phasing in climate dynamics (Fassbender et al., 2022; Nevison et al., 2016). Wind magnitude has a seasonal cycle, which additionally makes it important to get the phasing, not just the mean, of the seasonal $p\text{CO}_2$ cycle correct in order to capture the annual air-sea CO₂ flux (Fassbender et al., 2022; Rodgers et al., 2023). $p\text{CO}_2$ correlation is also the most robust element of the seasonal cycle to test in the Southern Ocean. While there are uncertainties in the magnitude of the seasonal $p\text{CO}_2$ amplitude and the CO₂ flux (i.e., differences in absolute $p\text{CO}_2$ magnitude between the SOCAT-only and SOCAT + SOCCOM products), results presented here for $p\text{CO}_2$ correlation are consistent regardless of which product is used. Additionally, one of the two mapping products we used (SOM-FFN) has been previously shown to accurately represent $p\text{CO}_2$ seasonality (Gloege et al., 2021).

The model PFAZ DIC amplitude biases described in the previous section equate to a range of -69% below to $+100\%$ above observed (mean $\pm$ standard deviation of $-16\% \pm 38\%$) and model DIC correlation biases equate to timing ranges from -44 days (early) to 34 days late relative to the observed DIC seasonal cycle (5 ± 19 days). Model SST amplitude biases are from -16% below to $+154\%$ above observed ($38 \pm 37\%$) and model SST timing ranges from -7 days (early) to 10 days late relative to the observed SST seasonal cycle (1 ± 4 days). Model TA amplitude biases range from -80% below to $+20\%$ above observed ($-46\% \pm 26\%$) and model TA timing ranges from -38 days (early) to 149 days late relative to the observed TA seasonal cycle (24 ± 40 days). We use these ranges of biases to set the bounds of the sensitivity testing conducted with our harmonic fit. We adjust each parameter (DIC, SST, TA) individually, using a consistent $\pm 50\%$ amplitude and ± 30 days to assess the ways in which biases in each parameter impacts the correlation between modeled $p\text{CO}_2$ and observed.

Symmetrically increasing the DIC amplitude by 50% (Figure 4a, dark blue dashed line) amplifies the seasonal cycle of low $p\text{CO}_2$ in the summer and high $p\text{CO}_2$ in the winter but has little impact on the $p\text{CO}_2$ timing ($\text{Corr}_{p\text{CO}_2} = 0.93$, Figures 4b and 4c, dark blue dashed line and circle). Conversely, symmetrically decreasing DIC amplitude by 50% (Figure 4a, light blue dashed line) produces a $p\text{CO}_2$ seasonal cycle that is out of phase ($\text{Corr}_{p\text{CO}_2} = -0.39$) as the DIC is rising too slowly in the winter and dropping too slowly in the summer, allowing seasonal temperature changes to dominate the $p\text{CO}_2$ impact.

What matters to $p\text{CO}_2$ timing is the relative rate of change of DIC and SST, and whether increases or decreases in DIC can outweigh the change in solubility induced by changing SST. In models the timing of seasonal cycles can obviously differ from reality. This means that the relative rate of change between DIC and SST can be impacted not only by the amplitude of the seasonal cycle, but by shifting the timing. Shifting the DIC timing by -30 days (i.e., winter maximum occurs 30 days earlier than observed, Figure 4d, light green dashed line) leads to an early peak in $p\text{CO}_2$ (Figure 4e) as DIC is now rising more rapidly relative to the increase in solubility caused by the SST drop from summer to winter. This lowers $\text{Corr}_{p\text{CO}_2}$ to 0.69 (Figure 4f, light green circle). Delaying the peak in DIC by 30 days (Figure 4d, dark green dashed line) has a much larger impact on $\text{Corr}_{p\text{CO}_2}$ (Figure 4e), shifting the seasonal $p\text{CO}_2$ to a $\text{Corr}_{p\text{CO}_2}$ of -0.28 (Figure 4f, dark green circle), as DIC rises more slowly relative to the increase in solubility from summer to winter such that $p\text{CO}_2$ initially drops instead of increasing into the winter. This is a crucial distinction from an analysis that separates $p\text{CO}_2$ impacts by only looking at the amplitude of the

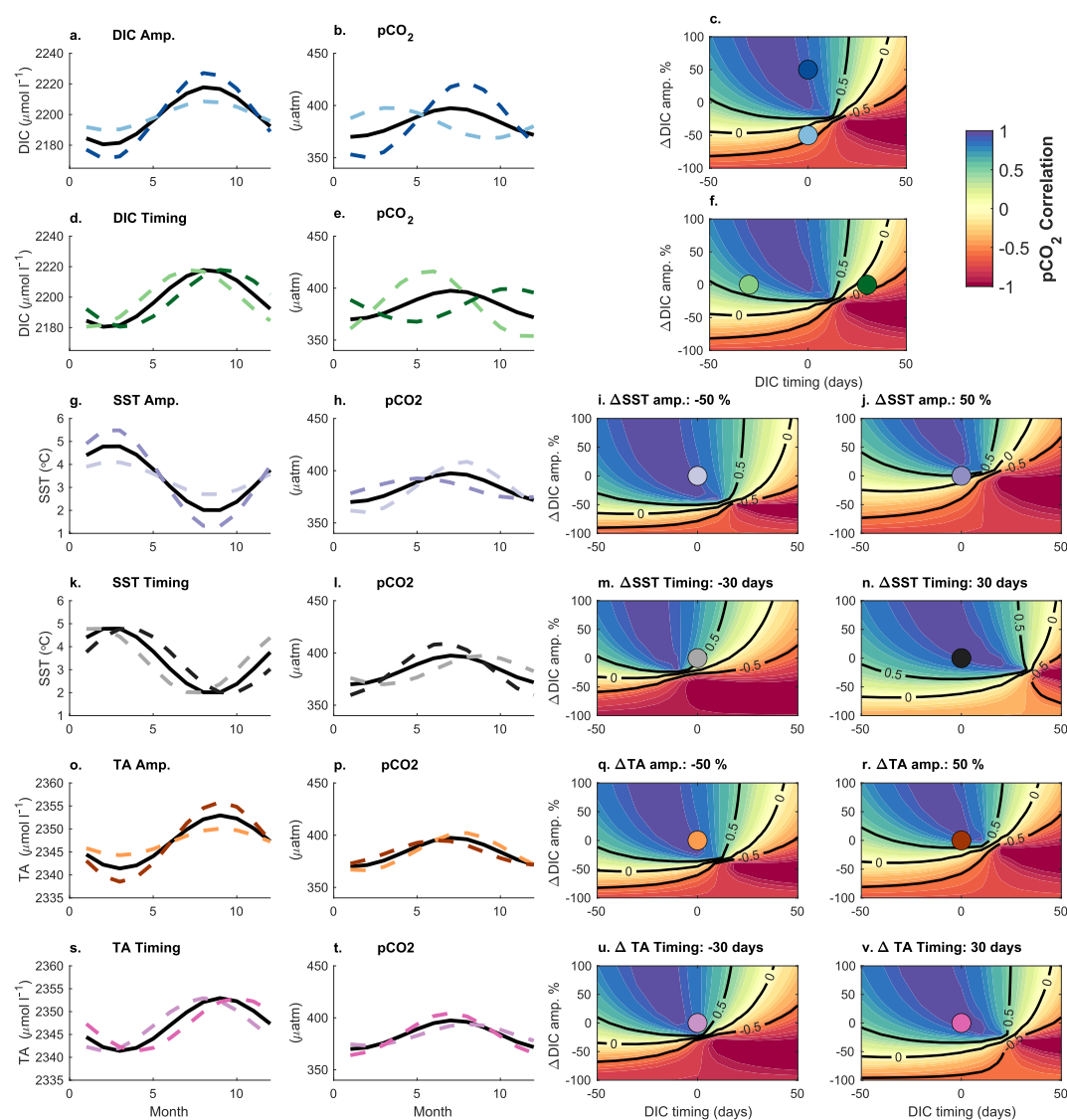


Figure 4. Harmonic fit response (see Sections 2.3 and 3.2) of $p\text{CO}_2$ to changes in dissolved inorganic carbon (DIC), sea surface temperature (SST), and TA seasonal cycles in PFAZ. Rows show individual parameter adjustments by $\pm 50\%$ for amplitude and ± 30 days for timing (variable changed indicated by the title of the leftmost panel). First and second columns: Observed mean seasonal cycles are shown in black lines. Dark/light shaded colored dashed lines show increased/decreased parameter (first column) and $p\text{CO}_2$ response (second column). Third and fourth columns: Contour plots of $p\text{CO}_2$ correlation to observations as a function of DIC amp. (%) and DIC timing (days). Contours in (c) and (f) are identical, while contour plots in rows 3–6 change according to the SST or TA change indicated in the title. Resulting $p\text{CO}_2$ correlations for each adjustment are indicated by dot location.

seasonal cycle, which would show no difference in the impact of DIC on $p\text{CO}_2$ between models with the same DIC amplitude but a shift in phase.

Shifting both DIC amplitude and timing can amplify the individual effects. For example, decreasing the DIC amplitude by 50% and DIC timing by +30 days (peak winter DIC occurring 30 days later than observations) produces a $\text{Corr}_{p\text{CO}_2}$ of -0.83 , a larger impact than either of those changes would individually incur. However, to limit the complexity of Figure 4 we only show changes in either the phase or amplitude of individual parameters.

Changing the SST amplitude can also increase or decrease the relative rate of change of DIC and SST. Decreasing the SST amplitude by 50% (Figure 4g, light purple dashed line) accentuates the impact of the seasonal DIC changes, leading to a moderate increase in the $p\text{CO}_2$ amplitude (Figure 4h) and little change in the correlation

($\text{Corr}_{p\text{CO}_2} = 0.95$, Figure 4i). However, most models exhibited an SST amplitude in excess of observed, which lowers the $p\text{CO}_2$ amplitude and shifts the peak $p\text{CO}_2$ to late summer/early fall, decreasing the correlation (Figure 4j, dark purple dot: $\text{Corr}_{p\text{CO}_2} = 0.48$ for a 50% increase in SST amplitude). Note that shifting the SST amplitude changes the correlation regions of the DIC amplitude versus DIC timing contour plots, which we display by showing two separate contour plots to accommodate this third dimension of change (Figure 4i vs. Figure 4j). Decreasing the SST amplitude expands the region of positive correlation, while increasing the amplitude decreases the area of positive correlation. Thus, for any combination of DIC amplitude and timing it is more likely that the $p\text{CO}_2$ will be out of phase if the SST amplitude is larger than observed.

Changes in SST timing can have an impact on $p\text{CO}_2$ correlation, with a shift of -30 days (early) decreasing $p\text{CO}_2$ correlation to 0.6 (Figure 4, light gray dashed line). Shifting SST timing late serves to reduce the change in solubility relative to the DIC change, so the $p\text{CO}_2$ seasonal cycle is accentuated, leading to a correlation of 0.96 (dark gray dashed line). We include changes in SST timing in Figure 4 for completeness, but since the mean model SST timing is only 1 ± 4 days later than observed, differences in SST timing do not have a significant impact on modeled $p\text{CO}_2$ in this region.

While the $p\text{CO}_2$ seasonal cycle is less sensitive to changes in TA amplitude or timing than equivalent changes in DIC or SST (Figure 4), differences between observed and modeled TA are, on average, larger than in DIC or SST (Figure 2) and can act to mask or accentuate the impact of biases in DIC and SST. Decreasing the TA amplitude by 50% (Figure 4o, light orange dashed line) improves $p\text{CO}_2$ correlations for most combinations of DIC amplitude and timing (Figure 4q). Increasing the TA amplitude has the opposite effect: when TA is higher in the winter than reality (Figure 4o, dark red dashed line) it brings $p\text{CO}_2$ down from what it would be in the winter (and up from the summer value) so that for all combinations of DIC amplitude and timing the $p\text{CO}_2$ correlation is slightly worse. TA timing that is 30 days early (-30 days, Figure 4s, light pink dashed line) decreases $\text{Corr}_{p\text{CO}_2}$ for all DIC amplitudes and timings (Figure 4u), while shifting the TA timing late by 30 days improves $\text{Corr}_{p\text{CO}_2}$ across all combinations of DIC amplitude and timing (Figure 4v).

To explore Earth system model and state estimate DIC, SST, and TA biases in the context of our harmonic fit, we overlaid model $p\text{CO}_2$ correlation on harmonic fit contour plots made with different representative biases (Figure 5). For example, the $\text{Corr}_{p\text{CO}_2}$ contours shown in Figure 5a are calculated with the observed TA and SST seasonal cycles and varying DIC amplitude (y-axis) and timing (x-axis), the same as shown in Figures 4c and 4f. Models that have an SST amplitude within $\pm 25\%$ and TA timing within ± 50 days of observed are overlain, with overall good agreement between $\text{Corr}_{p\text{CO}_2}$ calculated from model $p\text{CO}_2$ and that predicted by perturbing the harmonic fit, though with a shift in $\text{Corr}_{p\text{CO}_2}$ relative to the harmonic fit that we will discuss below. Moving down in the subplots represents an increase in the SST amplitude range (Figures 5c and 5d: 25%–75% SST amplitude; Figure 5e: 75%–175%) while moving to the right represents an increase in the TA timing shift to 50–150 days (Figures 5b and 5d). TA amplitude is also important, but to reduce the number of different plots shown here we do not separate models by TA amplitude in Figure 5 but include that dimension in Figures S16 and S17 in Supporting Information S1.

These plots indicate that, to first order, this simplified seasonal cycle captures the non-linear interactions that produce the seasonal $p\text{CO}_2$ cycle in models. For instance, viewed in this way it becomes clear that although both MPI-ESM1-2-LR(6) and ACCESS-ESM1-5(6) have similar, close to correct DIC amplitudes (Figure 5c), the late DIC timing of +21 days for ACCESS-ESM1-5(6) results in a $\text{Corr}_{p\text{CO}_2}$ of -0.97 , while MPI-ESM1-2-LR(6) retains a positive correlation ($\text{Corr}_{p\text{CO}_2} = 0.44$) despite a slightly larger, but early, timing shift of -27 days (Figure 5c). The importance of whether a timing shift is early or late is lost if we only consider the DIC correlation, as both models have a DIC correlation of ~ 0.9 (Table 2). Both of these models have an SST amplitude that is larger than observed, which also worsens the $\text{Corr}_{p\text{CO}_2}$ (i.e., Figure 5c vs. Figure 5a), though MPI-ESM1-2-LR(6) also has a dampened TA amplitude, which helps to improve its $\text{Corr}_{p\text{CO}_2}$ (Figure S16 in Supporting Information S1). CMCC-CESM, MRI-ESM1, and GFDL-ESM2M all have an even larger SST seasonal amplitude (Figure 5e). This drives each model's $p\text{CO}_2$ out of phase, even though their DIC amplitude and timing are relatively close to observations and would yield a positive $\text{Corr}_{p\text{CO}_2}$ with a correct SST amplitude (Figure 5a).

Conversely, NorESM1-ME, CMCC-ESM2(6), and AWI-ESM-1-RECoM(6) all have low DIC amplitude that would normally shift their $p\text{CO}_2$ out of phase (Figure 5b). However, their TA timing is shifted significantly later than observed (moving them into the right-hand column of subplots), which brings their $p\text{CO}_2$ into phase, giving a

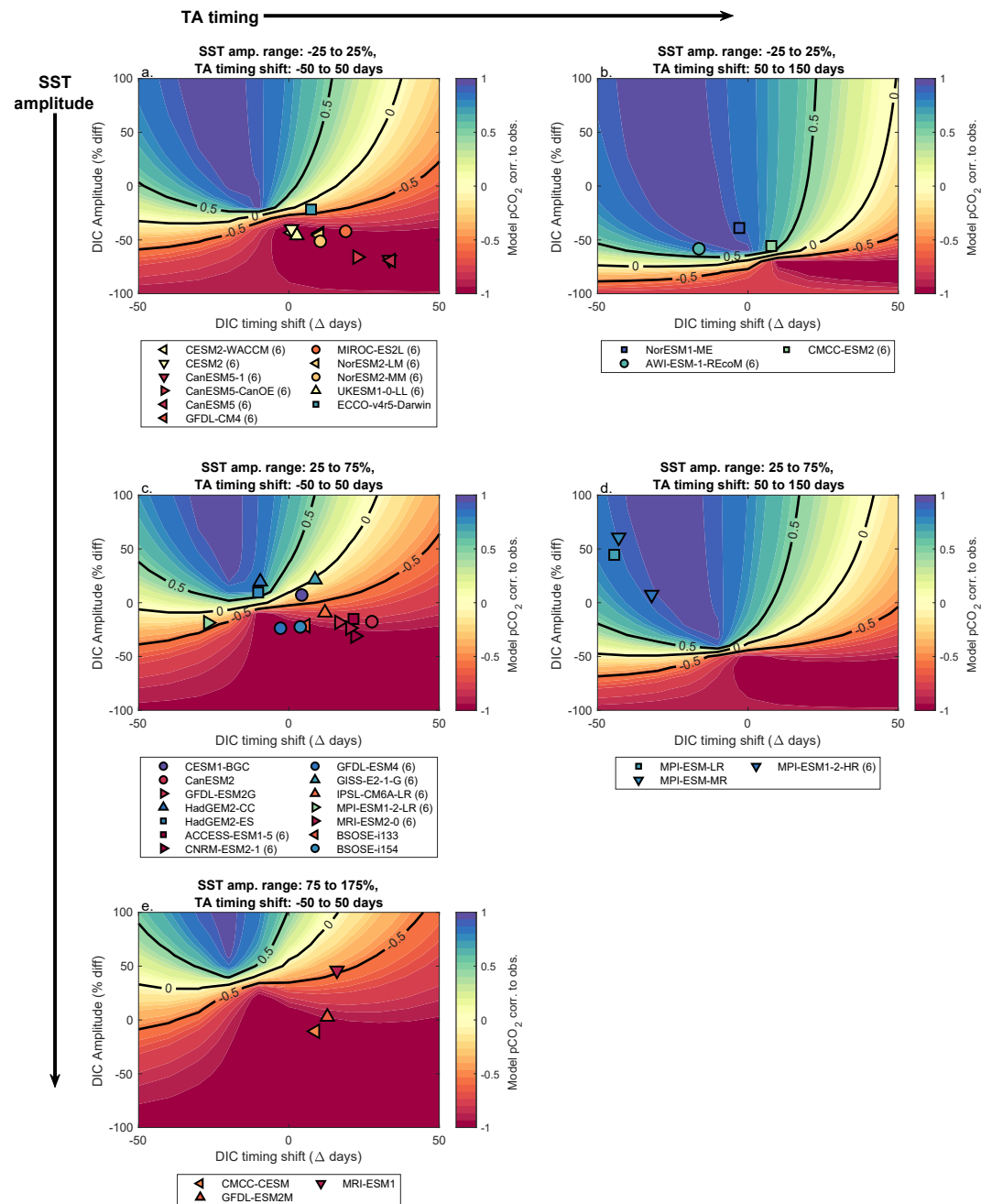


Figure 5. Harmonic fit $p\text{CO}_2$ correlation sensitivity with model-observational product $p\text{CO}_2$ correlation overlaid (symbols). Each panel shows harmonic fit $p\text{CO}_2$ correlation as a function of dissolved inorganic carbon (DIC) amplitude and DIC timing shift. Sea surface temperature (SST) amplitude range increases by row, while TA timing shift range increases by column. Top row: 0% SST amplitude (-25% to $+25\%$), middle: $+50\%$ ($+25\%$ to $+75\%$), bottom: 125% ($+75\%$ to $+175\%$). Left column: 0 day TA timing shift (-50 to $+50$ days). Right column: $+100$ day TA timing shift ($+50$ to $+150$ days). Models are sorted between panels according to their SST amplitude and TA timing. Model symbols are colored for their $p\text{CO}_2$ correlation based on the Taylor diagrams (Figure 2). Some differences between model $p\text{CO}_2$ correlation and that calculated for the harmonic fit are due to differences in TA amplitude, which are shown in Figures S16 and S17 in Supporting Information S1.

seemingly correct seasonal cycle of $p\text{CO}_2$ but for the wrong mechanistic reasons. Reduced TA amplitude has a similar effect, improving $\text{Corr}_{p\text{CO}_2}$ across all combinations of DIC amplitude and timing. For example, all of the models that fall in Figure 5a (SST amplitude between $\pm 25\%$ of observed, TA timing between ± 50 days of observed) have TA amplitudes that are smaller than observed. Figures S16 and S17 in Supporting Information S1

we add the dimension of TA amplitude ranges (-90 to -30% , -30% to 30% , 30% – 90%) with most models falling in the -90 to -30% range. This results in an improved correlation relative to what we expect based solely on DIC amplitude and timing and explains why the model symbols in Figures 5a and 5c tend to show a slightly better $\text{Corr}_{p\text{CO}_2}$ than the harmonic fit-derived contours underneath. For the range of TA amplitude and timings found in the models evaluated, TA timing shifts have a bigger impact on $\text{Corr}_{p\text{CO}_2}$, but more models have a reduced TA amplitude than a shift in TA timing, so both act to compensate biases in the DIC seasonal cycle.

3.3. Adjustments Needed to Bring Model $p\text{CO}_2$ Back in Phase With Observations

In Section 3.2 we showed that model biases in $p\text{CO}_2$ timing can be primarily explained by combinations of biases in DIC amplitude, DIC timing, and SST amplitude, with important contributions from biases in TA amplitude and TA timing. Here we examine which corrections to the mean model seasonal cycles of SST, DIC, and TA in the PFAZ are necessary to bring the $p\text{CO}_2$ seasonal cycle for each model back in phase. It is important to recognize that biases in model mean state can also impact the buffer capacity of seawater and therefore the sensitivity of $p\text{CO}_2$ to seasonal drivers. However, initial tests indicated that these effects do not impact our conclusions and are outside the scope of the present study.

As described in the methods, we conduct this sensitivity analysis using harmonic fits to the mean seasonal cycle. For models with seasonal cycles that are very different than the expected sinusoidal cycle, the $\text{Corr}_{p\text{CO}_2}$ determined from the harmonic fit (Table 3) can differ from that determined from the original seasonal $p\text{CO}_2$ (Table 2). In almost all cases these differences are small, but in the case of AWI-ESM-1-REcoM (6), the mean PFAZ has a $p\text{CO}_2$ seasonal cycle with two distinct peaks, one in summer when $p\text{CO}_2$ should be low and a slightly larger one in the winter when it is expected (Figure S41 in Supporting Information S1). The $\text{Corr}_{p\text{CO}_2}$ shown in Figure 2 and Table 2 is 0.6, but the harmonic fit, which produces a single peak in between the original two peaks, gives a $\text{Corr}_{p\text{CO}_2}$ of 0. This illustrates the limitations of this simple approach for testing a non-linear system, but we believe the benefits of a simply conducted and readily interpretable approach outweigh inclusion of multiple harmonics or a different, more complicated analysis.

The impacts of adjusting DIC, SST, and TA to correct for observed differences between mean model seasonal cycles and observed are shown for each model individually in Figures S29–S59 in Supporting Information S1 and summarized in Table 3 and Figure 6. Of the 34 models/state estimates with the required DIC, SST, TA, and SSS output available necessary to perform this analysis, 11 had a correlation ≥ 0.7 with no adjustments (Table 3). Importantly, of those 11, there are 7 that fall below the 0.7 correlation threshold when TA amplitude and/or phase are corrected, indicating that seasonal biases in TA were compensating for other biases in these seemingly skillful models (MPI-ESM-LR, MPI-ESM-MR, NorESM1-ME, GFDL-ESM4(6), MPI-ESM1-2-HR(6), BSOSE-i154, and ECCO-v4r5-Darwin). Similar compensating TA and DIC biases are often found in the annual mean magnitude of climate models (Hinrichs et al., 2023). In fact, correcting both model TA phase and amplitude only improves the $p\text{CO}_2$ correlation of 13 out of 34 models, while worsening the other 21. Only 4 models have good skill in $p\text{CO}_2$ correlation and remain so through all adjustments (CESM1-BGC, HadGEM2-CC, and HadGEM2-ES from CMIP5, and GISS-E2-1-G (6) from CMIP6, though HadGEM2-CC has a double peak in $p\text{CO}_2$ due to a broad and sustained DIC maximum that begins early in fall, Figure S34 in Supporting Information S1).

Correcting the DIC phase alone is the most consistent single variable for improving model $\text{Corr}_{p\text{CO}_2}$, improving 32 of the 34 models, though without bringing any additional models across the 0.7 correlation threshold (Table 3). Correcting DIC amplitude alone improves 26 of 34 models, with 18/34 (53%) now above 0.7 correlation, but worsens eight, with two models (MPI-ESM-LR and MPI-ESM-MR) dropping below 0.7 correlation as a result of the correction. Correcting both DIC phase and amplitude yields a total of 29/34 (85%) models above the 0.7 correlation threshold. Overall, nine models can be brought above a 0.7 $p\text{CO}_2$ correlation by adjusting the DIC amplitude alone (AWI-ESM-1-REcoM(6), CESM2-WACCM(6), CESM2(6), CMCC-ESM2(6), GFDL-CM4(6), MPI-ESM1-2-LR(6), NorESM2-LM(6), NorESM2-MM(6), and UKESM1-0-LL(6)), while nine others require both the DIC phase and amplitude to be corrected together (CanESM2, GFDL-ESM2G, CNRM-ESM2-1(6), CanESM5-1(6), CanESM5-CanOE(6), CanESM5(6), MIROC-ES2L(6), MRI-ESM2-0(6), and BSOSE-i133).

Improving the SST amplitude alone improves 25 of 34 models, with 15/34 (44%) total now above 0.7 $p\text{CO}_2$ correlation, though nine models are worsened. We do not show the impact of adjusting SST timing, as our calculations indicated that no model had an SST phase difference large enough to make an impact. Four models

Table 3
Response of $p\text{CO}_2$ Correlation to Adjustments of Model Dissolved Inorganic Carbon, Sea Surface Temperature, and Total Alkalinity Amplitude and Timing

Seasonal cycle $p\text{CO}_2$ correlations ^a to observed $p\text{CO}_2$, listed by parameter(s) adjusted										
Model name	Orig. (unadjusted)	DIC timing	DIC amp.	DIC timing, amp.	SST amp.	TA timing	TA amp.	TA timing, amp.	DIC timing, amp. SST timing, amp.	DIC, SST, TA
CMIP5										
CESM1-BGC	1.0	1.0	1.0	0.9	0.9	1.0	1.0	1.0	1.0	1.0
CMCC-CESM	−0.6	−0.3	−0.6	−0.0	1.0	−0.7	−0.6	−0.7	1.0	1.0
CanESM2	−0.8	0.6	−0.3	1.0	−0.1	−0.8	−1.0	−1.0	0.9	1.0
GFDL-ESM2G	−0.9	−0.0	−0.4	0.8	0.5	−0.9	−0.6	−0.8	0.9	1.0
GFDL-ESM2M	−0.9	0.1	−0.9	−0.0	0.7	−1.0	−0.4	−0.8	0.9	1.0
HadGEM2-CC	0.9	1.0	0.8	0.9	1.0	0.9	0.9	0.8	1.0	1.0
HadGEM2-ES	0.9	1.0	0.8	0.9	1.0	0.9	0.8	0.7	1.0	1.0
MPI-ESM-LR	0.7	0.9	0.5	1.0	0.8	0.6	0.7	0.6	1.0	1.0
MPI-ESM-MR	0.8	0.9	0.5	1.0	0.9	0.7	0.7	0.6	1.0	1.0
MRI-ESM1	−0.9	0.1	−0.9	−0.6	0.5	−0.9	−0.9	−0.9	1.0	1.0
NorESM1-ME	0.9	1.0	1.0	1.0	0.9	0.4	0.9	0.4	1.0	1.0
CMIP6										
ACCESS-ESM1-5 (6)	−1.0	−0.3	−0.9	0.2	−0.3	−1.0	−1.0	−1.0	1.0	1.0
AWI-ESM-1-REcoM (6)	0.0	−0.3	1.0	0.9	0.4	−0.2	0.7	−0.3	0.9	1.0
CESM2-WACCM (6)	−0.0	−0.0	1.0	1.0	0.2	−0.2	−0.2	−0.3	1.0	1.0
CESM2 (6)	0.0	0.0	1.0	1.0	0.4	−0.2	−0.1	−0.3	1.0	1.0
CMCC-ESM2 (6)	0.4	0.5	0.8	0.9	0.4	−0.4	0.8	−0.5	0.9	1.0
CNRM-ESM2-1 (6)	−0.9	−0.3	−0.5	0.8	−0.2	−1.0	−0.8	−0.9	0.9	1.0
CanESM5-1 (6)	−0.8	−0.4	0.1	0.9	−0.8	−0.8	−0.8	−0.8	0.9	1.0
CanESM5-CanOE (6)	−0.8	−0.3	0.4	0.9	−0.8	−0.8	−0.7	−0.7	0.9	1.0
CanESM5 (6)	−0.9	−0.4	0.0	0.9	−0.8	−0.8	−0.8	−0.7	0.9	1.0
GFDL-CM4 (6)	−0.8	−0.3	0.7	1.0	−0.8	−0.8	−0.6	−0.7	0.9	1.0
GFDL-ESM4 (6)	0.8	0.8	1.0	1.0	1.0	0.7	0.3	−0.1	0.9	1.0
GISS-E2-1-G (6)	0.9	1.0	0.9	0.9	0.8	0.8	0.9	0.9	1.0	1.0
IPSL-CM6A-LR (6)	−0.7	−0.1	−0.8	0.2	0.6	−0.6	−0.8	−0.6	1.0	1.0
MIROC-ES2L (6)	−0.6	0.3	0.6	1.0	0.7	−0.9	−0.2	−0.7	0.9	1.0
MPI-ESM1-2-HR (6)	0.8	0.9	0.8	1.0	0.9	0.7	0.8	0.6	0.9	1.0
MPI-ESM1-2-LR (6)	0.4	0.6	0.7	1.0	0.8	0.3	0.4	0.1	0.9	1.0
MRI-ESM2-0 (6)	−1.0	0.2	0.0	0.9	−0.1	−1.0	−0.9	−0.9	1.0	1.0
NorESM2-LM (6)	−0.5	0.1	0.7	0.9	−0.2	−0.8	−0.3	−0.6	0.9	1.0
NorESM2-MM (6)	−0.5	0.1	0.7	0.9	−0.6	−0.8	−0.3	−0.6	0.9	1.0
UKESM1-0-LL (6)	0.1	0.2	0.9	0.9	−0.1	0.2	−0.3	−0.2	0.9	1.0
State Estimates										
BSOSE-1133	−0.7	−0.9	0.5	0.8	0.4	0.3	−1.0	−0.3	0.8	1.0

Table 3
Continued

Model name	Seasonal cycle $p\text{CO}_2$ correlations ^a to observed $p\text{CO}_2$, listed by parameter(s) adjusted									DIC, SST, TA
	Orig. (unadjusted)	DIC timing	DIC amp.	DIC timing, amp.	SST amp.	TA timing	TA amp.	TA timing, amp.	DIC timing, amp. SST timing, amp.	
BSOSE-i154	0.8	1.0	0.8	0.9	0.8	0.9	−0.6	−0.3	0.9	1.0
ECCO-v4r5-Darwin	0.7	1.0	0.7	0.9	0.7	0.8	−0.9	0.5	0.9	1.0
# correlation ≥ 0.7 (%)	11 (32%)	11 (32%)	18 (53%)	29 (85%)	15 (44%)	9 (26%)	10 (29%)	4 (12%)	34 (100%)	34 (100%)
# improved		32	26	32	25	9	15	13	33	34
# worsened		2	8	2	9	25	19	21	1	0

^aCorrelations ≥ 0.7 are shown in bold text. Note that the $\text{Corr}_{p\text{CO}_2}$ in all columns are calculated from the harmonic fits to each parameter/observation-based product, not the original parameters. This includes the “Orig.” column, which therefore does not exactly match the $\text{Corr}_{p\text{CO}_2}$ in Table 2. Cells are shaded from red ($\text{Corr}_{p\text{CO}_2} = -1$) to blue ($\text{Corr}_{p\text{CO}_2} = 1$).

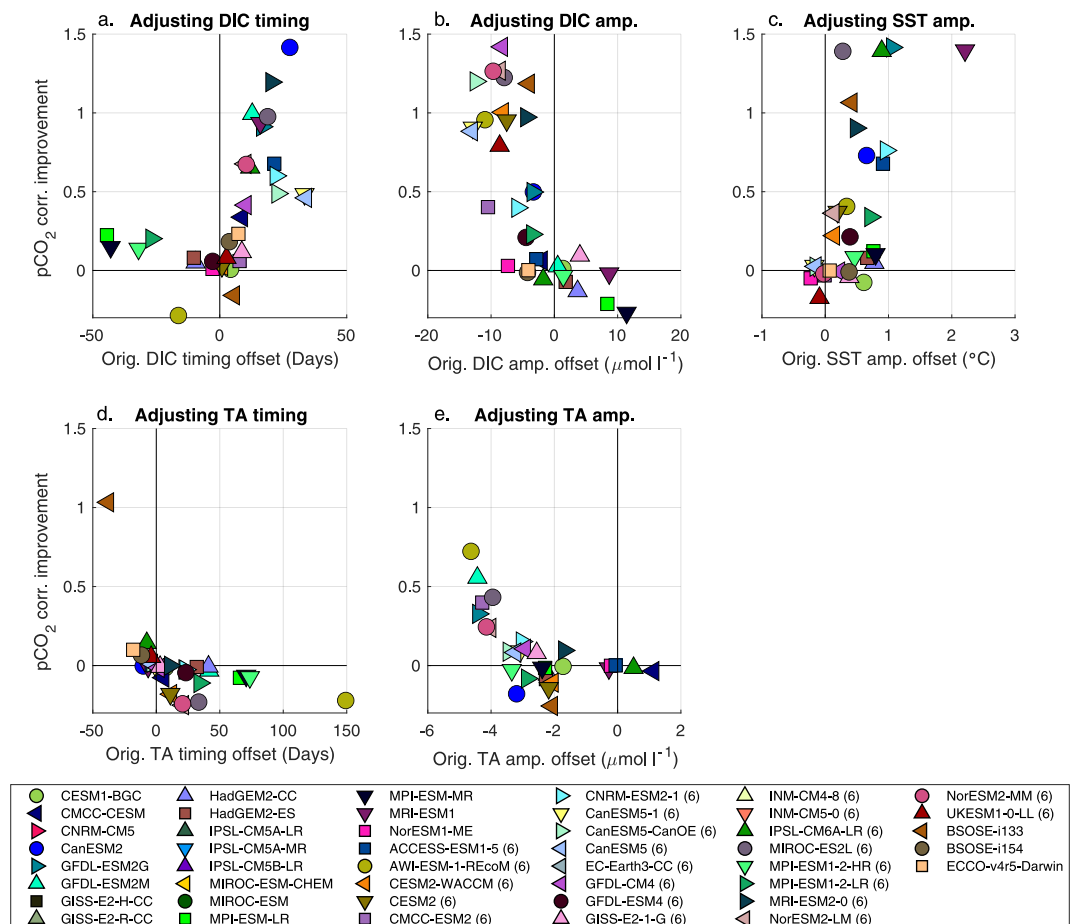


Figure 6. Impact on $p\text{CO}_2$ correlation of adjusting for model biases in (a) dissolved inorganic carbon (DIC) timing, (b) DIC amplitude, (c) sea surface temperature amplitude, (d), total alkalinity (TA) timing, and (e) TA amplitude. For each panel the x-axis is the original model harmonic fit compared to observations and the y-axis is the change in the model $p\text{CO}_2$ correlation due to adjusting that parameter alone. Results for all adjustments are in Table 3 and Figures S29–S59 in Supporting Information S1.

could be corrected by solely adjusting SST amplitude (CMCC-CESM, GFDL-ESM2M, MIROC-ES2L, and MPI-ESM1-2-LR (6)), while two needed both SST amplitude and DIC phase and/or amplitude corrected to cross the 0.7 $p\text{CO}_2$ correlation threshold (MRI-ESM1 and ACCESS-ESM1-5(6)). Correcting both SST amplitude and DIC phase and amplitude brings all 34 models above a $\text{Corr}_{p\text{CO}_2}$ of 0.7. Correction for TA in addition to DIC and SST can have additional improvements, but are not generally necessary to bring all models above a $p\text{CO}_2$ correlation of 0.7.

We included two iterations of the same state estimate, BSOSE, to give an example of how model development can fix the seasonal cycle of $p\text{CO}_2$, though not necessarily for the correct reasons. The later release, BSOSE-i154, has a MLD pattern that is more similar to observations than BSOSE-i133 (Figure S15 in Supporting Information S1) and a $p\text{CO}_2$ seasonal cycle that has an initial correlation of 0.8, compared to the BSOSE-i133 correlation of -0.7 . When looking at the seasonal cycles of SST, TA, and DIC for BSOSE-i133 (Figure S58 in Supporting Information S1) versus BSOSE-i154 (Figure S59 in Supporting Information S1) it appears that the largest change is in the TA cycle (Table 2). In this instance, the TA amplitude in BSOSE-i154 is much reduced relative to BSOSE-i133 and to observations, which spuriously brings the model $p\text{CO}_2$ into better correlation with observational products by suppressing $p\text{CO}_2$ in the summer and enhancing it in the winter.

We summarize the impact of model biases in DIC, SST, and TA by plotting the response of the $p\text{CO}_2$ correlation to correcting each individual offset against the magnitude of the original offset (Figure 6). Correcting models with late (+) DIC timing has a much larger improvement on the $p\text{CO}_2$ correlation than an early offset with similar magnitude (Figure 6a). As explained by using the harmonic fit perturbations in Figure 4, shifting the DIC timing earlier enhances the natural $p\text{CO}_2$ cycle, as the rate of change in DIC outpaces the counteracting change in temperature and therefore solubility. Similarly, correcting DIC amplitude that is too small can yield a significant improvement in $p\text{CO}_2$, while correcting DIC amplitudes that are too large worsens the model $p\text{CO}_2$ correlation to observational products (Figure 6b). Modeled SST amplitudes are, as described earlier, mostly greater than observed, leading to a rate of change in SST that changes the solubility too fast relative to changes in DIC; correcting these too-large SST amplitudes results in significant improvements to model $p\text{CO}_2$ timing (Figure 6c).

TA timing offsets do not usually have a large impact, with one notable exception in BSOSE-i133, which was 38 days early, correction of which improves the $p\text{CO}_2$ correlation by 1. TA timing in AWI-ESM-1-REcoM(6), on the other hand, was 150 days late, correction of which slightly worsened the model $\text{Corr}_{p\text{CO}_2}$ (Figure 6d). Correcting for the common problem of a too-small TA amplitude has a range of outcomes for the model $\text{Corr}_{p\text{CO}_2}$, depending on the TA timing and whether it was originally compensating for biases in DIC (Figure 6e).

3.4. Relationships of Model SST, DIC, and $p\text{CO}_2$ to Other Physical and Biological Properties

Section 3.2 demonstrated how the $p\text{CO}_2$ seasonal cycle is sensitive to relatively small biases in the amplitude and timing of SST, DIC, and TA and their non-linear interactions with $p\text{CO}_2$. In Section 3.3 we then identified common model biases in these parameters, explained how those biases resulted in changes in the timing of model $p\text{CO}_2$, and the impact of making different adjustments to bring model seasonal cycles of SST, DIC, and/or TA in line with observed. To better understand some of the underlying processes that either separately or jointly influence model biases in DIC, SST, TA, and therefore $p\text{CO}_2$, here we examine relationships between these properties and related model parameters. We recognize that correcting model biases is not straightforward, but explore these relationships to help demonstrate which model properties seem most related to the biases described above.

As previously discussed, biases in the magnitude of SST amplitude alone can be responsible for out-of-phase model $p\text{CO}_2$. Model SST amplitude is strongly correlated to MLD in summer months, with R^2 values above 0.6 from December through June (Figure 7). By comparing the relationship between MLD in each month and the magnitude of the seasonal SST amplitude for each model, we are deriving overall relationships present across most/all models. This correlation decreases during the progression to winter, reaching a minimum in August–October, when mixed layers are deepest. On average, the models assessed have shallower mixed layers and greater SST amplitude than observed. Shallow summertime mixed layers allow the same magnitude of surface heat flux to drive greater surface temperatures, due to the decreased thermal inertia of a shallow mixed layer (Ciasto & Thompson, 2008). Previous work has demonstrated that models have biases in surface radiative fluxes related to the transmission of radiation through clouds, with too much shortwave radiation in the summer and too

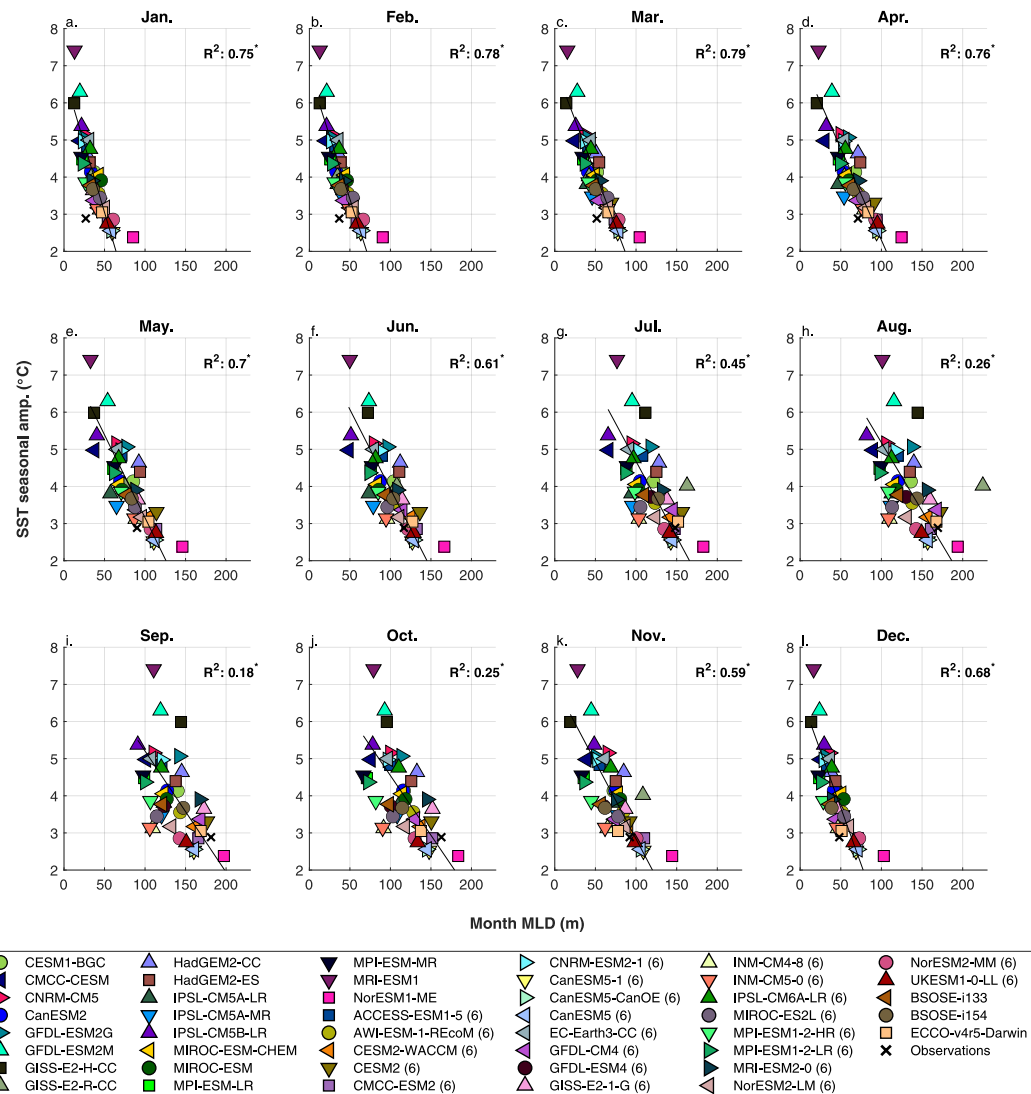


Figure 7. Relationship between sea surface temperature (SST) seasonal amplitude and mixed layer depth (MLD) by month. SST seasonal amplitude has a strong correlation with MLD from the late spring to early fall (R^2 of 0.6 and above from December to June), with shallow MLDs associated with greater seasonal SST amplitudes and deeper MLDs associated with smaller seasonal SST amplitudes. Regressions with a slope significantly different than 0 at a 95% confidence level are bold and marked with an “*.”

little longwave radiation in the winter (Cerovečki et al., 2022; Schneider & Reusch, 2016). Here, the same magnitude bias in MLD seems to have a much greater impact on SST amplitude if the MLD is biased too shallow than if the MLD is biased too deep (i.e., models with shallow MLDs tend to fall above the linear fit, Figure 7). In the winter, deeper mixed layers will have a greater thermal inertia due to the larger volume of water in contact with the atmosphere, limiting the temperature bias associated with an equivalently large deep MLD bias relative to a shallow-biased summer MLD.

Warm summertime MLs will drive up the $p\text{CO}_2$ for any given DIC. The shallow MLs that result in high summer SSTs also result in low summertime DIC, despite low wind speeds (Jones et al., 2014). DIC and SST amplitudes are positively correlated (Figure 8a). Interestingly, the observed DIC and SST amplitude (black “x”) falls above the regression between model DIC and SST amplitude, indicating that most models with the correct SST amplitude have a too-small DIC amplitude, and those with the correct DIC amplitude tend to have a too-large SST amplitude. Rodgers et al. (2023) found that reanalysis-forced global ocean models tended to also have a smaller amplitude seasonal DIC cycle than observed, which may indicate that this problem is not limited to Earth system

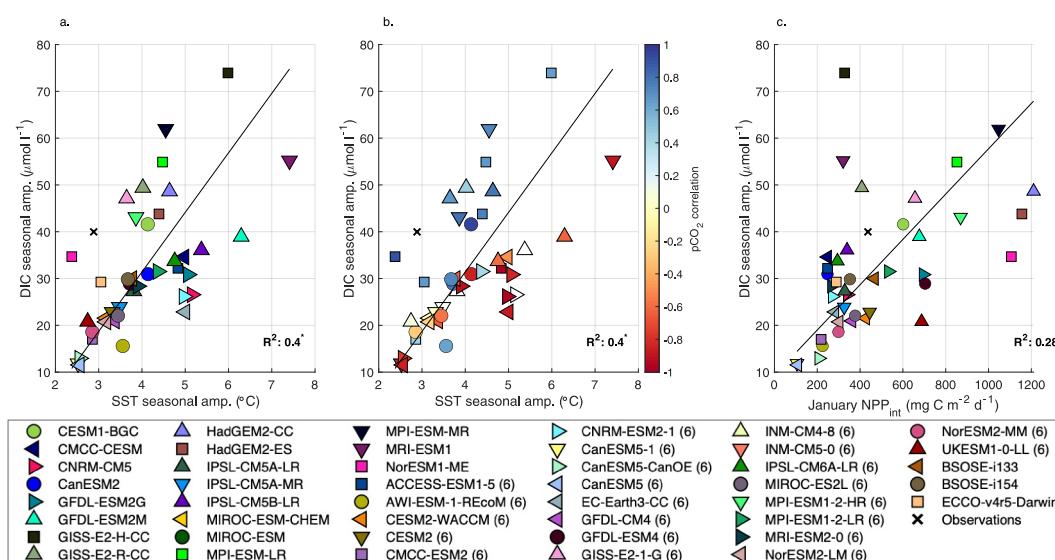


Figure 8. Relationship between dissolved inorganic carbon (DIC) seasonal amplitude, sea surface temperature (SST) amplitude, and January NPP_{int}. (a) DIC seasonal amplitude is correlated to SST seasonal amplitude ($R^2 = 0.4$). (b) When DIC amplitude is high relative to SST amplitude (symbols falling above the regression line), model $p\text{CO}_2$ correlation to observation-based $p\text{CO}_2$ tends to be positive (blue shaded symbols) and if SST amplitude is high relative to DIC the $p\text{CO}_2$ correlation is usually < 0 . (c) DIC seasonal amplitude correlation to January NPP_{int} ($R^2 = 0.28$). Regressions with a slope significantly different than 0 at a 95% confidence level are bold and marked with an “*”.

models. Both Rodgers et al. (2023) and Hauck, Gregor, et al. (2023) found that GOBMs' mean Southern Ocean SST seasonality and trend matched observations well, though they did not explore the seasonal SST in detail.

In Figure 7, we showed that shallow summertime mixed layers were associated with increased SST amplitude. Similarly, shallow MLDs are correlated with higher DIC amplitude, though the relationship explains less of the DIC amplitude differences between models than it does for SST (Figure S9 in Supporting Information S1). This is likely due to a combination of shallow summer mixed layers decreasing the residence time of DIC, allowing for a greater impact on DIC concentration for a given air-sea gas flux, and shallow mixed layers enabling light-limited phytoplankton to bloom, drawing down DIC. Nutrient draw-down could also be impacted, but the excess surface nutrients in the Southern Ocean make it unlikely that macro-nutrient limitation will play a factor. The lower correlation between DIC amplitude and monthly MLDs is also due in part to the outlier of NorESM1-ME. NorESM1-ME has consistently deep MLDs that are also evident in the annual mean (Figure S13 in Supporting Information S1) and subsequently a very low SST amplitude (Figure 7) but without a correspondingly small DIC amplitude. The overall relationships between DIC or SST amplitude and MLD can be summarized by plotting each against the ratio of the maximum monthly MLD to the minimum ($\text{MLD}_{\text{max}}/\text{MLD}_{\text{min}}$, Figures 9a and 9c). Differences in $\text{MLD}_{\text{max}}/\text{MLD}_{\text{min}}$ explain 55% of the variance of the DIC seasonal amplitude and 54% of the SST seasonal amplitude. We use the ratio of MLD_{max} to MLD_{min} rather than MLD_{min} alone, because the importance of a summertime (or wintertime) ML bias is due to both the magnitude of the bias relative to the season's MLD and relative to the overall range in MLD.

Unlike SST, DIC is also controlled by primary production. The phase (timing) of the DIC seasonal cycle is strongly correlated to the amplitude of NPP_{int} in the models ($R^2 = 0.64$, Figure 9b). Shallow summertime mixed layers mean that the DIC concentration can be more easily changed, so the timing of the DIC minima and the overall seasonal cycle is set by the strength of summertime productivity. This is supported by the relationship between the DIC seasonal amplitude and January NPP_{int}, with higher primary production associated with a greater seasonal amplitude of DIC ($R^2 = 0.28$, Figure 8). Some outliers exist and lower the regression fit, such as GISS-E2-H-CC, which has a very high DIC seasonal amplitude despite a relatively low January NPP_{int}.

The non-linearity of the carbonate system means that $p\text{CO}_2$ timing is only weakly correlated with most variables, despite the strong sensitivity demonstrated in Sections 3.2 and 3.3. However, overall $p\text{CO}_2$ normalized RMSE and amplitude are related to parameters in ways that support the controls already identified. For instance, $p\text{CO}_2$

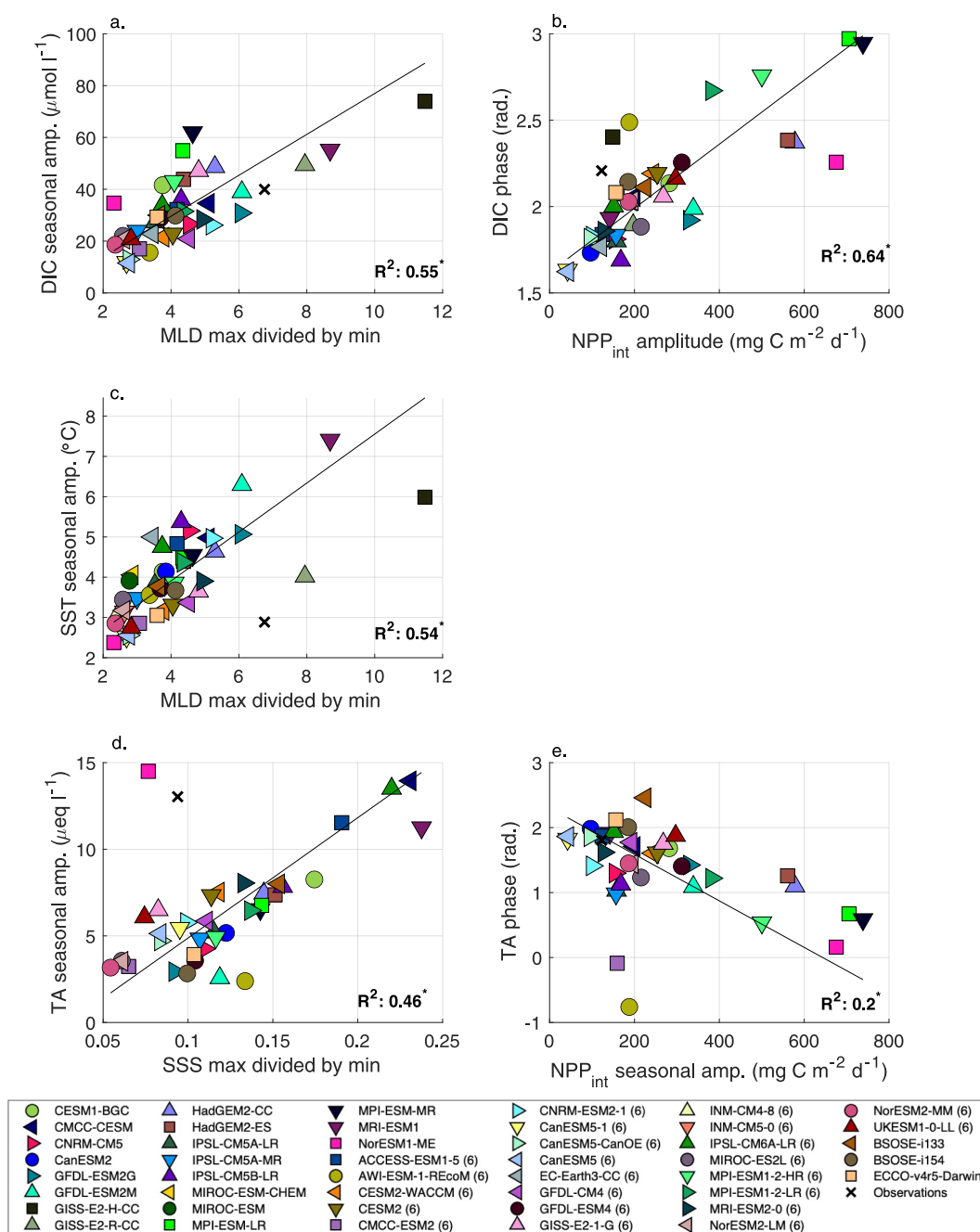


Figure 9. Relationships between model carbonate system parameters and other relevant model parameters. Dissolved inorganic carbon (DIC) seasonal amplitude is related to (a) the maximum monthly mixed layer depth divided by the minimum ($\text{MLD}_{\text{max}}/\text{MLD}_{\text{min}}$, see individual monthly relationships in Figure S9 in Supporting Information S1) and (b) DIC phase correlates to the integrated net primary production (NPP_{int}) amplitude. (c) sea surface temperature seasonal amplitude is also correlated to $\text{MLD}_{\text{max}}/\text{MLD}_{\text{min}}$ (Figure 7). (d) TA seasonal amplitude is related to the maximum sea surface salinity divided by the minimum and (e) TA phase is related to NPP_{int} amplitude. Regressions with a slope significantly different than 0 at a 95% confidence level are bold and marked with an “*.”

normalized RMSE is strongly related to model DIC correlation with observations (Figure 10a, $R^2 = 0.63$). This further supports our harmonic fit experiments where we demonstrated how changes in DIC timing can worsen $p\text{CO}_2$ fit in different ways. When DIC peaks too early in the winter, $p\text{CO}_2$ correlation remains good, but the amplitude increases beyond what is observed, and when DIC peaks too late, $p\text{CO}_2$ quickly goes out of phase.

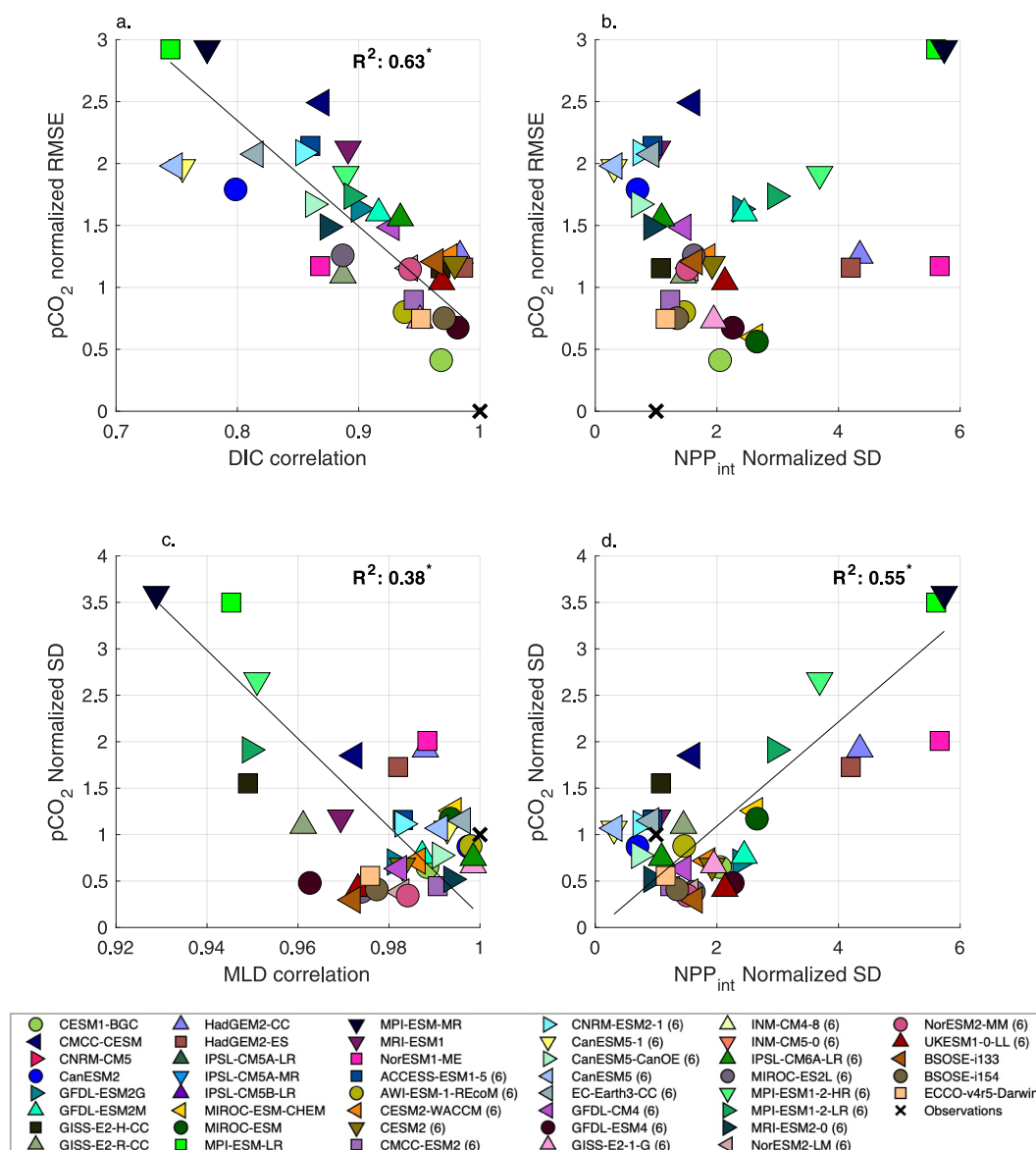


Figure 10. Model seasonal $p\text{CO}_2$ normalized RMSE versus (a) dissolved inorganic carbon correlation to observations and (b) NPP_{int} normalized SD as well as $p\text{CO}_2$ normalized SD (model $p\text{CO}_2$ SD divided by observed) versus (c) mixed layer depth correlation to observations and (d) integrated net primary production (NPP_{int}) normalized SD. Black “x”s denote observed quantities. Regressions (black lines) with a slope significantly different than 0 at a 95% confidence level are bold and marked with an “*.”

The $p\text{CO}_2$ normalized SD is also related to NPP_{int} normalized SD (Figure 10d, $R^2 = 0.55$), with higher SD NPP_{int} generally found with higher $p\text{CO}_2$ SD. This is likely due to high carbon drawdown by NPP_{int} in the spring/summer lowering DIC, strengthening the $p\text{CO}_2$ seasonal cycle of low in the summer and high in the winter (Figures 8 and 9), increasing the $p\text{CO}_2$ amplitude. The lowest $p\text{CO}_2$ error is found with normalized NPP_{int} SD values 2x that of the NPP_{int} average found from the satellite derived products (Figure 10b), which may indicate that too high NPP_{int} can artificially create a large DIC seasonal amplitude (Lerner et al., 2021), yielding a good $p\text{CO}_2$ correlation. Overly strong NPP has been previously shown to result in too low surface $p\text{CO}_2$ in the summer (Kessler & Tjiputra, 2016). Other possible alternatives are that the NPP products are biased or that the organic matter produced by excessive NPP is rapidly respired (McClish et al., 2026), thus not leading to strong DIC/ $p\text{CO}_2$ drawdown. However, even a seasonal drawdown of carbon that does not result in carbon export has been shown to have significant impacts on the annual air-sea flux (Huang et al., 2023).

$p\text{CO}_2$ normalized SD, or relative amplitude, is also related to MLD correlation to observations (Figure 10c, $R^2 = 0.38$). There is little variation in MLD correlation, but for those models with comparatively low MLD correlation (<0.95), the normalized $p\text{CO}_2$ SD is much higher than observed. MPI-ESM-LR and MPI-ESM-MR have poor MLD correlations, poor DIC correlations, and high NPP_{int} ratios, indicating systemic issues across multiple parameters.

In summary, model SST and DIC amplitudes are both related to the MLD biases, with shallow MLDs related to greater SST and DIC seasonal amplitudes. The impacts of larger SST and DIC seasonal amplitudes on $p\text{CO}_2$ oppose each other, with the net effect on model $\text{Corr}_{p\text{CO}_2}$ dependent on which wins out (Figure 8). Correcting model MLD will likely improve both SST and DIC amplitudes (Danek & Hauck, 2026). Whether that leads to subsequent improvements in $p\text{CO}_2$ correlation depends on the existence of other model biases. NPP_{int} is strongly related to the timing of DIC and TA seasonal cycles. While we did not find that SSS biases on their own were a significant contributor to $\text{Corr}_{p\text{CO}_2}$ biases, TA amplitude is related to the SSS seasonal maximum to minimum ratio (Figure 9d). These overlapping interactions between physical and biological drivers illustrate the unsurprising difficulty in determining how to improve a model's representation of the carbonate system. However, from this analysis it seems clear that modeled MLD is a good target for addressing DIC and SST amplitude biases, while the timing of DIC and TA are most sensitive to the strength of NPP_{int} .

4. Conclusions

Correctly simulating the timing of $p\text{CO}_2$ requires the right balance of seasonal cycles of SST, DIC, and TA, and therefore can be used as a test of mechanistic fidelity in models. Significant insight has been gained through past work from a determination of whether $p\text{CO}_2$ in a region is “thermally” or “non-thermally” driven (e.g., Takahashi et al., 2002) by leveraging the seasonal amplitude of SST and/or DIC. We now have sufficient confidence in the timing of the seasonal $p\text{CO}_2$ to explore how shifts in SST, DIC, and TA timing impact $p\text{CO}_2$ timing, both alone and in conjunction with model biases in amplitude.

We focus on the critical Southern Ocean region between the Subantarctic Front and 62°S due to its significance to the regional and global carbon cycle and the difficulty that models have in reproducing the seasonal carbon cycle and air-sea fluxes in this region. In doing so we describe how biases in DIC, SST, and TA timing and amplitude impact the $p\text{CO}_2$ seasonal cycle.

Only 11 out of the 45 models (19 CMIP5, 23 CMIP6, and 3 state estimate simulations) assessed in this study have a relatively good correlation with observed Polar Frontal Antarctic Zone seasonal CO_2 fluxes. Using a harmonic fit of the mean seasonal cycles of DIC, SST, TA, and SSS in the PFAZ, we explore the sensitivity of $p\text{CO}_2$ to changes in both the amplitude and timing of carbonate system parameters. Only four out of the 34 models for which we could conduct the seasonal adjustment exercise appeared to have the correct $p\text{CO}_2$ correlation with no major compensating biases, in other words, the right answer for the right reasons. The seasonal cycle of $p\text{CO}_2$ is very sensitive to the relative amplitudes of the DIC and SST seasonal cycles as well as the timing of DIC seasonality. $p\text{CO}_2$ is also very sensitive to SST timing changes, but all models correctly represented the SST seasonal timing so that does not play an important role in $p\text{CO}_2$ biases. Seasonal $p\text{CO}_2$ is less sensitive to TA amplitude and timing biases relative to DIC and SST, but the biases in the TA cycle are much larger, which makes TA issues an important consideration for understanding out of phase model $p\text{CO}_2$.

DIC and SST amplitude biases are clearly related to differences in monthly MLD or seasonal amplitude. The shallower the summer ML or the greater the ratio of the maximum to minimum MLs, the greater the amplitude of SST and DIC seasonal cycles. The models explored here have, on average, both shallower summertime and wintertime MLs than observed. While the MLD seasonal timing is usually quite close to observed, very small biases in the MLD correlation can also greatly impact the $p\text{CO}_2$ amplitude. Together, these sensitive relationships make an accurate seasonal MLD a critical concern when determining how to improve model $p\text{CO}_2$ correlation. Indeed, it may be that addressing MLD biases is the clearest path toward improving model representation of ocean biogeochemistry in this region.

Differences in the amplitude of integrated NPP are directly related to the timing of DIC and TA. It is not clear if satellite-based estimates of NPP are providing an accurate target for models, but they provide a good starting point that can potentially be combined in the future with other broad tracers of seasonal biological production, such as

Acknowledgments

SMB was originally funded under National Science Foundation SOCCOM Grant OPP-1936222. SMB has additionally received support from NASA Carbon Monitoring System Program (80NSSC23K1231), NASA Carbon Cycle Science (80NSSC22K0156), SOCCOM3 (NSF OCE-2332379), and received support from Schmidt Sciences, LLC (InMOS). MRM acknowledges funding from NSF Grant OPP-2149501. IC acknowledges funding from NSF Grant OPP-2409690. PL is supported by the European Union's Horizon Europe Research and Innovation programme under Grant 101199028 (TRICUSO) and Grant 101137682 (AI4PEX), and received support from Schmidt Sciences, LLC (OBVI InMOS). AJF was supported by PMEL. JH and CD acknowledge funding from the European Union under the ERC-2022-STG OceanPeak (Grant 101077209). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Council. Neither the European Union nor the granting authority can be held responsible for them. This work used resources of the Deutsches Klimarechenzentrum (DKRZ) granted by its Scientific Steering Committee (WLA) under project ID ba1103. ARG gratefully acknowledges support from NSF Awards 1946578, 2110258, and 2332379 and from NOAA Award NA24OARX431C0057. This publication is partially funded by the Cooperative Institute for Climate, Ocean, & Ecosystem Studies (CICOES) under NOAA Cooperative Agreements NA20OAR4320271 and NA25OARX432C0012. Contribution No. 2026-1570. This is PMEL contribution number 5799. This is SOEST contribution 12198. This work would not have begun without the mentorship and guidance of Jorge Sarmiento and we owe much of our understanding of ocean biogeochemistry, especially in the Southern Ocean, to his work and insight. He will be missed. Taylor calculations and harmonic fit code from Kathie Kelly. Thanks to T. E. Bushinsky and L. M. Bushinsky for their insights and suggestions (e.g., "you should fix that"). We would also like to thank the careful and thoughtful reviews provided by J. Orr, P. Mongwe, and G. McKinley.

upper ocean nutrient cycles or oxygen saturation anomalies, to evaluate the strength of seasonal biological production.

The approach that we developed to test non-linear interactions driving seasonal $p\text{CO}_2$ using a harmonic fit of the mean carbonate system can be used to understand how different biogeographical regions (e.g., Fay & McKinley, 2014) of the ocean are sensitive to different carbonate system parameter biases. When considering how to evaluate different regions, it is important to select areas in both models and observations that have similar seasonal cycles in $p\text{CO}_2$. For instance, the SAF location varies between models, but is commonly the boundary between net CO_2 out-gassing and net uptake, while the Polar Front bisects the region of high winter $p\text{CO}_2$. Understanding whether models match observed seasonal cycles, and not just mean annual uptake, hinges on correctly selecting these regions. Correctly identifying biases in models and relating those biases to specific features and processes is important for understanding how the ocean may respond to climate change, which has already been shown to increase the magnitude of the $p\text{CO}_2$ seasonal cycle (Fassbender et al., 2018) and has the potential to significantly alter the amplitude of SST, DIC, and TA (Orr et al., 2022) and shift the relative timing of their cycles. Accurately representing the drivers of the $p\text{CO}_2$ seasonal cycle in models will be critical to projecting air-sea flux representations of future climate.

For now, this analysis provides the sensitivity of Southern Ocean $p\text{CO}_2$ seasonal cycles to physical and chemical drivers with the goal of using these observational constraints to improve future global climate models. We note that three of the four models with correct $p\text{CO}_2$ timing and no compensating biases are CMIP5 models, which may indicate that model development efforts may need to explicitly target the seasonal CO_2 cycle to improve in this area. Model development focusing foremost on accurately capturing the seasonal cycle of the MLD is a necessary first step to a better representation of the seasonal carbonate system in the Southern Ocean.

Conflict of Interest

The authors declare no conflicts of interest relevant to this study.

Availability Statement

Float data were collected and made freely available by the Southern Ocean Carbon and Climate Observations and Modeling (SOCCOM, Johnson et al., 2017; Sarmiento et al., 2023) Project funded by the National Science Foundation, Division of Polar Programs (NSF PLR-1425989 and OPP-1936222 and OCE-2332379), supplemented by NASA, and by the International Argo Program and the NOAA programs that contribute to it (<http://www.argo.ucsd.edu>, <https://www.ocean-ops.org/board>, Roemmich & Gilson, 2009). The Argo Program is part of the Global Ocean Observing System. The SOCCOM April 2023 data snapshot is from Riser et al. (2023). The Surface Ocean CO_2 Atlas (SOCAT) is an international effort, endorsed by the SCOR Infrastructural Project International Ocean Carbon Coordination Project (IOCCP) and the Surface Ocean Lower-Atmosphere Study (SOLAS), to deliver a uniform, quality-controlled surface ocean CO_2 database. The many researchers and funding agencies responsible for the collection of data and quality control are thanked for their contributions to SOCAT. SOCATv2023 data can be accessed from Bakker et al. (2023). NOAA OI SST v2 are from (<https://psl.noaa.gov/data/gridded/data.noaa.oisst.v2.html>, Reynolds et al., 2002). Model output citations are listed in Table 1 and primarily downloaded from the Earth System Grid Federation (<https://aims2.llnl.gov/search/>) with the following exceptions: CESM1-BGC output downloaded from <https://www.earthsystemgrid.org>, BSOSE output downloaded from: [sose.ucsd.edu](https://www.earthsystemgrid.org), AWI-ESM-1-RECoM (CMIP 6) from C. Danek, ECCO-v4r5 from D. Carroll and D. Menemenlis. Mixed layer depth is calculated from Roemmich-Gilson gridded Argo data (Roemmich & Gilson, 2009). NPP_{int} estimates from the Ocean Productivity site (<http://orca.science.oregonstate.edu/index.php>, Behrenfeld & Falkowski, 1997; Behrenfeld et al., 2005; Westberry et al., 2008; Silsbe et al., 2016). All code and intermediate data files needed to reproduce the figures and analysis, CO_2 flux, and $p\text{CO}_2$ data products used in this work are published in Zenodo (<https://zenodo.org/records/17517686>, Bushinsky et al., 2026).

References

- Adachi, Y., Yukimoto, S., Deushi, M., Obata, A., Nakano, H., Tanaka, T. Y., et al. (2013). Basic performance of a new Earth system model of the Meteorological Research Institute (MRI-ESM1). *Papers in Meteorology and Geophysics*, 64, 1–19. <https://doi.org/10.2467/mripapers.64.1>
- Anav, A., Friedlingstein, P., Kidston, M., Bopp, L., Ciais, P., Cox, P., et al. (2013). Evaluating the land and ocean components of the global carbon cycle in the CMIP5 Earth system models. *Journal of Climate*, 26(18), 6801–6843. <https://doi.org/10.1175/JCLI-D-12-00417.1>

- Bakker, D. C. E., Alin, S. R., Bates, N., Becker, M., Feely, R. A., Gkritzalis, T., et al. (2023). *Surface Ocean CO₂ Atlas database version 2023 (SOCATv2023)* (NCEI Accession 0278913). NOAA National Centers for Environmental Information. <https://doi.org/10.25921/R7XA-BT92>
- 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 fCO₂ data in version 3 of the Surface Ocean CO₂ Atlas (SOCAT). *Earth System Science Data*, 8(2), 383–413. <https://doi.org/10.5194/essd-8-383-2016>
- Behrenfeld, M. J., Boss, E., Siegel, D. A., & Shea, D. M. (2005). Carbon-based ocean productivity and phytoplankton physiology from space. *Global Biogeochemical Cycles*, 19(1), GB1006. <https://doi.org/10.1029/2004GB002299>
- Behrenfeld, M. J., & Falkowski, P. G. (1997). Photosynthetic rates derived from satellite-based chlorophyll concentration. *Limnology and Oceanography*, 42(1), 1–20. <https://doi.org/10.4319/lo.1997.42.1.0001>
- Boucher, O., Servonnat, J., Albright, A. L., Aumont, O., Balkanski, Y., Bastrikov, V., et al. (2020). Presentation and evaluation of the IPSL-CM6A-LR climate model. *Journal of Advances in Modeling Earth Systems*, 12(7), e2019MS002010. <https://doi.org/10.1029/2019MS002010>
- Bushinsky, S. M., Arteaga Quintero, L. A., Fassbender, A., Hauck, J., Mazloff, M., Ceroveck, I., et al. (2026). Data and code for: Evaluation of timing and amplitude biases in the controls of simulated Southern Ocean pCO₂ [Dataset]. *Zenodo*. <https://doi.org/10.5281/zenodo.17517686>
- Bushinsky, S. M., Landschützer, P., Rödenbeck, 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>
- Bushinsky, S. M., Nachod, Z., Fassbender, A. J., Tamsitt, V., Takeshita, Y., & Williams, N. (2025). Offset between profiling float and shipboard oxygen observations at depth impacts bias on float pH and derived pCO₂. *Global Biogeochemical Cycles*, 39(5), e2024GB008185. <https://doi.org/10.1029/2024GB008185>
- Canadell, J. G., Monteiro, P. M., Costa, M. H., Da Cunha, L. C., Cox, P. M., Eliseev, A. V., et al. (2021). Global carbon and other biogeochemical cycles and feedbacks. In *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* (1st ed.). Cambridge University Press. <https://doi.org/10.1017/9781009157896.007>
- Carranza, M. M., Long, M. C., Di Luca, A., Fassbender, A. J., Johnson, K. S., Takeshita, Y., et al. (2024). Extratropical storms induce carbon outgassing over the Southern Ocean. *npj Climate and Atmospheric Science*, 7(1), 106. <https://doi.org/10.1038/s41612-024-00657-7>
- Carroll, D., Menemenlis, D., Adkins, J. F., Bowman, K. W., Brix, H., Dutkiewicz, S., et al. (2020). The ECCO-Darwin data-assimilative global ocean biogeochemistry model: Estimates of seasonal to multi-decadal surface ocean pCO₂ and air-sea CO₂ flux. *Journal of Advances in Modeling Earth Systems*, 12(10), e2019MS001888. <https://doi.org/10.1029/2019MS001888>
- Ceroveck, I., Sun, R., Bromwich, D. H., Zou, X., Mazloff, M. R., & Wang, S.-H. (2022). Impact of downward longwave radiative deficits on Antarctic sea-ice extent predictability during the sea ice growth period. *Environmental Research Letters*, 17(8), 084008. <https://doi.org/10.1088/1748-9326/ac7d66>
- Christian, J. R., Denman, K. L., Hayashida, H., Holdsworth, A. M., Lee, W. G., Riche, O. G. J., et al. (2022). Ocean biogeochemistry in the Canadian Earth System Model version 5.0.3: CanESM5 and CanESM5-CanOE. *Geoscientific Model Development*, 15(11), 4393–4424. <https://doi.org/10.5194/gmd-15-4393-2022>
- Ciasto, L. M., & Thompson, D. W. J. (2008). Observations of large-scale ocean–atmosphere interaction in the Southern Hemisphere. *Journal of Climate*, 21(6), 1244–1259. <https://doi.org/10.1175/JCLI11809.1>
- Danek, C., Gürses, z., Gierz, P., Andrés-Martínez, M., & Hauck, J. (2023). AWI AWI-ESMIREcoM model output prepared for CMIP6 CMIP. Earth System Grid Federation. <https://doi.org/10.22033/ESGF/CMIP6.17722>
- Danek, C., & Hauck, J. (2024). Discrepancies in temporal pCO₂ variability from Earth System Models and pCO₂-products related to high-latitude mixed layer dynamics and equatorial upwelling. *Research Square*. <https://doi.org/10.21203/rs.3.rs-3851521/v1>
- Danek, C., & Hauck, J. (2026). Contrasting high-latitude mixed layer depth trends in Earth System Models and data products and their impact on temporal pCO₂ variability. *Climate Dynamics*, 64(5), 211. <https://doi.org/10.1007/s00382-026-08109-z>
- 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>
- DeVries, T. (2014). The oceanic anthropogenic CO₂ sink: Storage, air-sea fluxes, and transports over the industrial era. *Global Biogeochemical Cycles*, 28(7), 631–647. <https://doi.org/10.1002/2013GB004739>
- Dickson, A. G. (1990). Standard potential of the reaction: AgCl(s) + 12H₂(g) = Ag(s) + HCl(aq), and the standard acidity constant of the ion HSO₄[−] in synthetic sea water from 273.15 to 318.15 K. *The Journal of Chemical Thermodynamics*, 22(2), 113–127. [https://doi.org/10.1016/0021-9614\(90\)90074-Z](https://doi.org/10.1016/0021-9614(90)90074-Z)
- Döscher, R., Acosta, M., Alessandri, A., Anthoni, P., Arsouze, T., Bergman, T., et al. (2022). The EC-Earth3 Earth system model for the Coupled Model Intercomparison Project 6. *Geoscientific Model Development*, 15(7), 2973–3020. <https://doi.org/10.5194/gmd-15-2973-2022>
- Dufresne, J.-L., Foujols, M.-A., Denvil, S., Caubel, A., Marti, O., Aumont, O., et al. (2013). Climate change projections using the IPSL-CM5 Earth System Model: From CMIP3 to CMIP5. *Climate Dynamics*, 40(9), 2123–2165. <https://doi.org/10.1007/s00382-012-1636-1>
- Dunne, J. P., Horowitz, L. W., Adcroft, A. J., Ginoux, P., Held, I. M., John, J. G., et al. (2020). The GFDL Earth System Model version 4.1 (GFDL-ESM 4.1): Overall coupled model description and simulation characteristics. *Journal of Advances in Modeling Earth Systems*, 12(11), e2019MS002015. <https://doi.org/10.1029/2019MS002015>
- Dunne, J. P., John, J. G., Adcroft, A. J., Griffies, S. M., Hallberg, R. W., Shevliakova, E., et al. (2012). GFDL's ESM2 global coupled climate–carbon Earth system models. Part I: Physical formulation and baseline simulation characteristics. *Journal of Climate*, 25(19), 6646–6665. <https://doi.org/10.1175/JCLI-D-11-00560.1>
- Dunne, J. P., John, J. G., Shevliakova, S., Stouffer, R. J., Krasting, J. P., Malyshev, S. L., et al. (2013). GFDL's ESM2 global coupled climate–carbon Earth system models. Part II: Carbon system formulation and baseline simulation characteristics. *Journal of Climate*, 26(7), 2247–2267. <https://doi.org/10.1175/jcli-d-12-00150.1>
- Fassbender, A. J., Rodgers, K. B., Palevsky, H. I., & Sabine, C. L. (2018). Seasonal asymmetry in the evolution of surface ocean pCO₂ and pH thermodynamic drivers and the influence on sea-air CO₂ flux. *Global Biogeochemical Cycles*, 32(10), 1476–1497. <https://doi.org/10.1029/2017GB005855>
- Fassbender, A. J., Schlunegger, S., Rodgers, K. B., & Dunne, J. P. (2022). Quantifying the role of seasonality in the marine carbon cycle feedback: An ESM2M case study. *Global Biogeochemical Cycles*, 36(6), e2021GB007018. <https://doi.org/10.1029/2021GB007018>
- Fay, A., Gregor, L., Landschützer, P., McKinley, G., 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>
- Fay, A., & 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>

- Forget, G., Campin, J.-M., Heimbach, P., Hill, C. N., Ponte, R. M., & Wunsch, C. (2015). ECCO version 4: An integrated framework for non-linear inverse modeling and global ocean state estimation. *Geoscientific Model Development*, 8(10), 3071–3104. <https://doi.org/10.5194/gmd-8-3071-2015>
- Friedlingstein, P., O'Sullivan, M., Jones, M. W., Andrew, R. M., Bakker, D. C. E., Hauck, J., et al. (2023). Global Carbon Budget 2023. *Earth System Science Data*, 15(12), 5301–5369. <https://doi.org/10.5194/essd-15-5301-2023>
- Fu, W., Moore, J. K., Primeau, F., Collier, N., Ogunro, O. O., Hoffman, F. M., & Randerson, J. T. (2022). Evaluation of ocean biogeochemistry and carbon cycling in CMIP Earth system models with the International Ocean Model Benchmarking (IOMB) software system. *Journal of Geophysical Research: Oceans*, 127(10), e2022JC018965. <https://doi.org/10.1029/2022JC018965>
- Fuss, S., Canadell, J. G., Peters, G. P., Tavoni, M., Andrew, R. M., Ciais, P., et al. (2014). Betting on negative emissions. *Nature Climate Change*, 4(10), 850–853. <https://doi.org/10.1038/nclimate2392>
- Gallego, M. A., Timmermann, A., Friedrich, T., & Zeebe, R. E. (2018). Drivers of future seasonal cycle changes in oceanic $p\text{CO}_2$. *Biogeosciences*, 15(17), 5315–5327. <https://doi.org/10.5194/bg-15-5315-2018>
- Giorgetta, M. A., Jungclaus, J., Reick, C. H., Legutke, S., Bader, J., Böttinger, M., et al. (2013). Climate and carbon cycle changes from 1850 to 2100 in MPI-ESM simulations for the Coupled Model Intercomparison Project phase 5. *Journal of Advances in Modeling Earth Systems*, 5(3), 572–597. <https://doi.org/10.1002/jame.20038>
- Gloag, L., McKinley, G. A., Landschützer, P., Fay, A. R., Frölicher, T. L., Fyfe, J. C., 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>
- Goris, N., Tjiputra, J. F., Olsen, A., Schwinger, J., Lauvset, S. K., & Jeansson, E. (2018). Constraining projection-based estimates of the future North Atlantic carbon uptake. *Journal of Climate*, 31(10), 3959–3978. <https://doi.org/10.1175/JCLI-D-17-0564.1>
- Gray, A. R., Johnson, K. S., Bushinsky, S. M., Riser, S. C., Russell, J. L., Talley, L. D., et al. (2018). Autonomous biogeochemical floats detect significant carbon dioxide outgassing in the high-latitude Southern Ocean. *Geophysical Research Letters*, 45(17), 9049–9057. <https://doi.org/10.1029/2018GL078013>
- Gruber, N., Landschützer, P., & Lovenduski, N. S. (2019). The variable Southern Ocean carbon sink. *Annual Review of Marine Science*, 11(1), 159–186. <https://doi.org/10.1146/annurev-marine-121916-063407>
- Hajima, T., Watanabe, M., Yamamoto, A., Tatebe, H., Noguchi, M. A., Abe, M., et al. (2020). Development of the MIROC-ES2L Earth system model and the evaluation of biogeochemical processes and feedbacks. *Geoscientific Model Development*, 13(5), 2197–2244. <https://doi.org/10.5194/gmd-13-2197-2020>
- Hauck, J., Gregor, L., Nissen, C., Patara, L., Hague, M., Mongwe, P., et al. (2023). The Southern Ocean carbon cycle 1985–2018: Mean, seasonal cycle, trends, and storage. *Global Biogeochemical Cycles*, 37(11), e2023GB007848. <https://doi.org/10.1029/2023GB007848>
- Hauck, J., Nissen, C., Landschützer, P., Rödenbeck, C., Bushinsky, S., & Olsen, A. (2023). Sparse observations induce large biases in estimates of the global ocean CO_2 sink: An ocean model subsampling experiment. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 381(2249), 20220063. <https://doi.org/10.1098/rsta.2022.0063>
- Hauck, J., & Völker, C. (2015). Rising atmospheric CO_2 leads to large impact of biology on Southern Ocean CO_2 uptake via changes of the Revelle factor. *Geophysical Research Letters*, 42(5), 1459–1464. <https://doi.org/10.1002/2015GL063070>
- Hauck, J., Völker, C., Wolf-Gladrow, D. A., Laufkötter, C., Vogt, M., Aumont, O., et al. (2015). On the Southern Ocean CO_2 uptake and the role of the biological carbon pump in the 21st century. *Global Biogeochemical Cycles*, 29(9), 1451–1470. <https://doi.org/10.1002/2015GB005140>
- Held, I. M., Guo, H., Adcroft, A., Dunne, J. P., Horowitz, L. W., Krasting, J., et al. (2019). Structure and performance of GFDL's CM4.0 climate model. *Journal of Advances in Modeling Earth Systems*, 11(11), 3691–3727. <https://doi.org/10.1029/2019MS001829>
- Hinrichs, C., Köhler, P., Völker, C., & Hauck, J. (2023). Alkalinity biases in CMIP6 Earth system models and implications for simulated CO_2 drawdown via artificial alkalinity enhancement. *Biogeosciences*, 20(18), 3717–3735. <https://doi.org/10.5194/bg-20-3717-2023>
- Huang, Y., Fassbender, A. J., & Bushinsky, S. M. (2023). Biogenic carbon pool production maintains the Southern Ocean carbon sink. *Proceedings of the National Academy of Sciences*, 120(18), e2217909120. <https://doi.org/10.1073/pnas.2217909120>
- Ilyina, T., Six, K. D., Segsneider, J., Maier-Reimer, E., Li, H., & Núñez-Riboni, I. (2013). Global ocean biogeochemistry model HAMOCC: Model architecture and performance as component of the MPI-Earth system model in different CMIP5 experimental realizations. *Journal of Advances in Modeling Earth Systems*, 5(2), 287–315. <https://doi.org/10.1029/2012MS000178>
- Johnson, K. S., Plant, J. N., Coletti, L. J., Jannasch, H. W., Sakamoto, C. M., Riser, S. C., et al. (2017). Biogeochemical sensor performance in the SOCCOM profiling float array. *Journal of Geophysical Research: Oceans*, 122(8), 6416–6436. <https://doi.org/10.1002/2017JC012838>
- Jones, D. C., Ito, T., Takano, Y., & Hsu, W.-C. (2014). Spatial and seasonal variability of the air-sea equilibration timescale of carbon dioxide. *Global Biogeochemical Cycles*, 28(11), 1163–1178. <https://doi.org/10.1002/2014GB004813>
- Kelley, M., Schmidt, G. A., Nazarenko, L. S., Bauer, S. E., Ruedy, R., Russell, G. L., et al. (2020). GISS-E2.1: Configurations and climatology. *Journal of Advances in Modeling Earth Systems*, 12(8), e2019MS002025. <https://doi.org/10.1029/2019MS002025>
- Kessler, A., & Tjiputra, J. (2016). The Southern Ocean as a constraint to reduce uncertainty in future ocean carbon sinks. *Earth System Dynamics*, 7(2), 295–312. <https://doi.org/10.5194/esd-7-295-2016>
- Key, R., Olsen, A., van Heuven, S., Lauvset, S., Velo, A., Lin, X., et al. (2015). *Global Ocean Data Analysis Project, Version 2 (GLODAPv2)*. ORNL/CDIAC-162, NDP-093. https://doi.org/10.3334/CDIAC/OTG.NDP093_GLODAPv2
- Landschützer, P., Gruber, N., Bakker, D. C. E., Schuster, U., Nakaoka, S., Payne, M. R., et al. (2013). A neural network-based estimate of the seasonal to inter-annual variability of the Atlantic Ocean carbon sink. *Biogeosciences*, 10(11), 7793–7815. <https://doi.org/10.5194/bg-10-7793-2013>
- Landschützer, P., Gruber, N., Bakker, D. C. E., Stemmler, I., & Six, K. D. (2018). Strengthening seasonal marine CO_2 variations due to increasing atmospheric CO_2 . *Nature Climate Change*, 8(2), 146–150. <https://doi.org/10.1038/s41558-017-0057-x>
- Lauderdale, J. M., Dutkiewicz, S., Williams, R. G., & Follows, M. J. (2016). Quantifying the drivers of ocean-atmosphere CO_2 fluxes. *Global Biogeochemical Cycles*, 30(7), 983–999. <https://doi.org/10.1002/2016GB005400>
- Lee, K., Kim, T.-W., Byrne, R. H., Millero, F. J., Feely, R. A., & Liu, Y.-M. (2010). The universal ratio of boron to chlorinity for the North Pacific and North Atlantic oceans. *Geochimica et Cosmochimica Acta*, 74(6), 1801–1811. <https://doi.org/10.1016/j.gca.2009.12.027>
- Lerner, P., Romanou, A., Kelley, M., Romanski, J., Ruedy, R., & Russell, G. (2021). Drivers of air-sea CO_2 flux seasonality and its long-term changes in the NASA-GISS model CMIP6 submission. *Journal of Advances in Modeling Earth Systems*, 13(2), e2019MS002028. <https://doi.org/10.1029/2019MS002028>
- Long, M. C., Lindsay, K., Peacock, S., Moore, J. K., & Doney, S. C. (2013). Twentieth-century oceanic carbon uptake and storage in CESM1 (BGC). *Journal of Climate*, 26(18), 6775–6800. <https://doi.org/10.1175/JCLI-D-12-00184.1>
- Lovato, T., Peano, D., Butenschön, M., Matera, S., Iovino, D., Scoccimarro, E., et al. (2022). CMIP6 simulations with the CMCC Earth system model (CMCC-ESM2). *Journal of Advances in Modeling Earth Systems*, 14(3), e2021MS002814. <https://doi.org/10.1029/2021MS002814>

- Lueker, T. J., Dickson, A. G., & Keeling, C. D. (2000). Ocean $p\text{CO}_2$ 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](https://doi.org/10.1016/S0304-4203(00)00022-0)
- Mackallah, C., Chamberlain, M. A., Law, R. M., Dix, M., Ziehn, T., Bi, D., et al. (2022). ACCESS datasets for CMIP6: Methodology and idealised experiments. *Journal of Southern Hemisphere Earth Systems Science*, 72(2), 93–116. <https://doi.org/10.1071/ES21031>
- Marti, O., Braconnot, P., Dufresne, J.-L., Bellier, J., Benshila, R., Bony, S., et al. (2010). Key features of the IPSL ocean atmosphere model and its sensitivity to atmospheric resolution. *Climate Dynamics*, 34(1), 1–26. <https://doi.org/10.1007/s00382-009-0640-6>
- Martin, G. M., Bellouin, N., Collins, W. J., Culverwell, I. D., Halloran, P. R., Hardiman, S. C., et al. (2011). The HadGEM2 family of Met Office Unified Model climate configurations. *Geoscientific Model Development*, 4(3), 723–757. <https://doi.org/10.5194/gmd-4-723-2011>
- Maurer, T. L., Plant, J. N., & Johnson, K. S. (2021). Delayed-mode quality control of oxygen, nitrate, and pH data on SOCCOM biogeochemical profiling floats. *Frontiers in Marine Science*, 8, 683207. <https://doi.org/10.3389/fmars.2021.683207>
- Mauritsen, T., Bader, J., Becker, T., Behrens, J., Bittner, M., Brokopf, R., et al. (2019). Developments in the MPI-M Earth System Model version 1.2 (MPI-ESM1.2) and its response to increasing CO_2 . *Journal of Advances in Modeling Earth Systems*, 11(4), 998–1038. <https://doi.org/10.1029/2018MS001400>
- McClish, S., Bushinsky, S., Briggs, N., Cerovečki, I., & Douglas, C. (2026). Seasonal re-entrainment of respired organic matter decouples Southern Ocean surface production and annual net community production. *AGU Advances*, 7(5), e2025AV002267. <https://doi.org/10.1029/2025AV002267>
- Meehl, G. A., Arblaster, J. M., Bates, S., Richter, J. H., Tebaldi, C., Gettelman, A., et al. (2020). Characteristics of future warmer base states in CESM2. *Earth and Space Science*, 7(9), e2020EA001296. <https://doi.org/10.1029/2020EA001296>
- Mikaloff Fletcher, S. E., Gruber, N., Jacobson, A. R., Gloor, M., Doney, S. C., Dutkiewicz, S., et al. (2007). Inverse estimates of the oceanic sources and sinks of natural CO_2 and the implied oceanic carbon transport. *Global Biogeochemical Cycles*, 21(1), GB1010. <https://doi.org/10.1029/2006GB002751>
- Mongwe, P., Gregor, L., Tjiputra, J., Hauck, J., Ito, T., Danek, C., et al. (2024). Projected poleward migration of the Southern Ocean CO_2 sink region under high emissions. *Communications Earth & Environment*, 5(1), 1–13. <https://doi.org/10.1038/s43247-024-01382-y>
- Mongwe, P., Vichi, M., & Monteiro, P. M. S. (2018). The seasonal cycle of $p\text{CO}_2$ and CO_2 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>
- Moore, J. K., Doney, S. C., & Lindsay, K. (2004). Upper ocean ecosystem dynamics and iron cycling in a global three-dimensional model. *Global Biogeochemical Cycles*, 18(4), GB4028. <https://doi.org/10.1029/2004GB002220>
- Nevison, C. D., Manizza, M., Keeling, R. F., Stephens, B. B., Bent, J. D., Dunne, J., et al. (2016). Evaluating CMIP5 ocean biogeochemistry and Southern Ocean carbon uptake using atmospheric potential oxygen: Present-day performance and future projection. *Geophysical Research Letters*, 43(5), 2077–2085. <https://doi.org/10.1002/2015GL067584>
- Nickford, S., Palter, J. B., Donohue, K., Fassbender, A. J., Gray, A. R., Long, J., et al. (2022). Autonomous wintertime observations of air-sea exchange in the Gulf Stream reveal a perfect storm for ocean CO_2 uptake. *Geophysical Research Letters*, 49(5), e2021GL096805. <https://doi.org/10.1029/2021GL096805>
- Olsen, A., Key, R. M., Van Heuven, S., Lauvset, S. K., Velo, A., Lin, X., et al. (2016). The Global Ocean Data Analysis Project version 2 (GLODAPv2)—An internally consistent data product for the world ocean. *Earth System Science Data*, 8(2), 297–323. <https://doi.org/10.5194/essd-8-297-2016>
- Orr, J. C., Kwiatkowski, L., & Pörtner, H.-O. (2022). Arctic Ocean annual high in $p\text{CO}_2$ could shift from winter to summer. *Nature*, 610(7930), 94–100. <https://doi.org/10.1038/s41586-022-05205-y>
- Palmer, J., & Totterdell, I. (2001). Production and export in a global ocean ecosystem model. *Deep Sea Research Part I: Oceanographic Research Papers*, 48(5), 1169–1198. [https://doi.org/10.1016/S0967-0637\(00\)00080-7](https://doi.org/10.1016/S0967-0637(00)00080-7)
- Perez, F. F., & Fraga, F. (1987). Association constant of fluoride and hydrogen ions in seawater. *Marine Chemistry*, 21(2), 161–168. [https://doi.org/10.1016/0304-4203\(87\)90036-3](https://doi.org/10.1016/0304-4203(87)90036-3)
- Peters, G. P., Andrew, R. M., Boden, T., Canadell, J. G., Ciais, P., Le Quéré, C., et al. (2013). The challenge to keep global warming below 2°C . *Nature Climate Change*, 3(1), 4–6. <https://doi.org/10.1038/nclimate1783>
- Prend, C. J., Hunt, J. M., Mazloff, M. R., Gille, S. T., & Talley, L. D. (2022). Controls on the boundary between thermally and non-thermally driven $p\text{CO}_2$ regimes in the South Pacific. *Geophysical Research Letters*, 49(9), 1–11. <https://doi.org/10.1029/2021GL095797>
- Reagan, J. R., Seidov, D., Wang, Z., Dukhovskoy, D., Boyer, T. P., Locarnini, R. A., et al. (2024). World Ocean Atlas 2023. In A. Mishonov (Ed.), *Salinity* (Vol. 2). NOAA Atlas NESDIS 90. <https://doi.org/10.25923/70qt-9574>
- 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](https://doi.org/10.1175/1520-0442(2002)015<1609:AIISAS>2.0.CO;2)
- Riser, S. C., Talley, L. D., Wijffels, S. E., Nicholson, D., Purkey, S., Takeshita, Y., et al. (2023). SOCCOM and GO-BGC float data—Snapshot 2023-04-26. In *Southern Ocean Carbon and Climate Observations and Modeling (SOCCOM) and Global Ocean Biogeochemistry (GO-BGC) Biogeochemical-Argo Float Data Archive*. UC San Diego Library Digital Collections. <https://doi.org/10.6075/J05B45XQ>
- 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 $p\text{CO}_2$ Mapping intercomparison (SOCOM). *Biogeosciences Discussions*, 12(16), 14049. <https://doi.org/10.5194/bgd-12-14049-2015>
- Rödenbeck, C., Keeling, R. F., Bakker, D. C., Metzl, N., Olsen, A., Sabine, C., & Heimann, M. (2013). Global surface-ocean $p\text{CO}_2$ and sea-air CO_2 flux variability from an observation-driven ocean mixed-layer scheme. *Ocean Science*, 9(2), 193–216. <https://doi.org/10.5194/os-9-193-2013>
- Rodgers, K. B., Sarmiento, J. L., Aumont, O., Crevoisier, C., de Boyer Montégut, C., & Metzl, N. (2008). A wintertime uptake window for anthropogenic CO_2 in the North Pacific. *Global Biogeochemical Cycles*, 22(2), GB2020. <https://doi.org/10.1029/2006GB002920>
- Rodgers, K. B., Schwinger, J., Fassbender, A. J., Landschützer, P., Yamaguchi, R., Frenzel, H., et al. (2023). Seasonal variability of the surface ocean carbon cycle: A synthesis. *Global Biogeochemical Cycles*, 37(9), e2023GB007798. <https://doi.org/10.1029/2023GB007798>
- Roemmich, D., & Gilson, J. (2009). The 2004–2008 mean and annual cycle of temperature, salinity, and steric height in the global ocean from the Argo Program. *Progress in Oceanography*, 82(2), 81–100. <https://doi.org/10.1016/j.pocan.2009.03.004>
- Romanou, A., Gregg, W., Romanski, J., Kelley, M., Bleck, R., Healy, R., et al. (2013). Natural air–sea flux of CO_2 in simulations of the NASA-GISS climate model: Sensitivity to the physical ocean model formulation. *Ocean Modelling*, 66, 26–44. <https://doi.org/10.1016/j.ocemod.2013.01.008>
- Rustogi, P., Landschützer, P., Brune, S., & Baehr, J. (2023). The impact of seasonality on the annual air-sea carbon flux and its interannual variability. *npj Climate and Atmospheric Science*, 6(1), 1–8. <https://doi.org/10.1038/s41612-023-00378-3>

- Sarmiento, J. L., Johnson, K. S., Arteaga, L. A., Bushinsky, S. M., Cullen, H. M., Gray, A. R., et al. (2023). The Southern Ocean carbon and climate observations and modeling (SOCCOM) project: A review. *Progress in Oceanography*, 219, 103130. <https://doi.org/10.1016/j.pocean.2023.103130>
- Schmidt, G. A., Kelley, M., Nazarenko, L., Ruedy, R., Russell, G. L., Aleinov, I., et al. (2014). Configuration and assessment of the GISS ModelE2 contributions to the CMIP5 archive. *Journal of Advances in Modeling Earth Systems*, 6(1), 141–184. <https://doi.org/10.1002/2013MS000265>
- Schneider, D. P., & Reusch, D. B. (2016). Antarctic and Southern Ocean surface temperatures in CMIP5 models in the context of the surface energy budget. *Journal of Climate*, 29(5), 1689–1716. <https://doi.org/10.1175/JCLI-D-15-0429.1>
- Schulzweida, U. (2023). CDO user guide [Dataset]. *Zenodo*. <https://doi.org/10.5281/ZENODO.10020800>
- S  f  rian, R., Bopp, L., Gehlen, M., Orr, J. C., Eth  , C., Cadule, P., et al. (2013). Skill assessment of three Earth system models with common marine biogeochemistry. *Climate Dynamics*, 40(9–10), 2549–2573. <https://doi.org/10.1007/s00382-012-1362-8>
- S  f  rian, R., Nabat, P., Michou, M., Saint-Martin, D., Voldoire, A., Colin, J., et al. (2019). Evaluation of CNRM Earth system model, CNRM-ESM2-1: Role of Earth system processes in present-day and future climate. *Journal of Advances in Modeling Earth Systems*, 11(12), 4182–4227. <https://doi.org/10.1029/2019MS001791>
- Seland, O., Bentsen, M., Oliivi  , D., Toniazzo, T., Gjermundsen, A., Graff, L. S., et al. (2020). Overview of the Norwegian Earth System Model (NorESM2) and key climate response of CMIP6 DECK, historical, and scenario simulations. *Geoscientific Model Development*, 13(12), 6165–6200. <https://doi.org/10.5194/gmd-13-6165-2020>
- Sellar, A. A., Jones, C. G., Mulcahy, J. P., Tang, Y., Yool, A., Wiltshire, A., et al. (2019). UKESM1: Description and evaluation of the U.K. Earth System Model. *Journal of Advances in Modeling Earth Systems*, 11(12), 4513–4558. <https://doi.org/10.1029/2019MS001739>
- Semmler, T., Danilov, S., Gierz, P., Goessling, H. F., Hegewald, J., Hinrichs, C., et al. (2020). Simulations for CMIP6 with the AWI climate model AWI-CM-1-1. *Journal of Advances in Modeling Earth Systems*, 12(9), e2019MS002009. <https://doi.org/10.1029/2019MS002009>
- Sigmund, M., Anstey, J., Arora, V., Digby, R., Gillett, N., Kharin, V., et al. (2023). Improvements in the Canadian Earth System Model (CanESM) through systematic model analysis: CanESM5.0 and CanESM5.1. *Geoscientific Model Development*, 16(22), 6553–6591. <https://doi.org/10.5194/gmd-16-6553-2023>
- Silsbe, G. M., Behrenfeld, M. J., Halsey, K. H., Milligan, A. J., & Westberry, T. K. (2016). The CAFE model: A net production model for global ocean phytoplankton. *Global Biogeochemical Cycles*, 30(12), 1756–1777. <https://doi.org/10.1002/2016GB005521>
- Swart, N. C., Cole, J. N. S., Kharin, V. V., Lazare, M., Scinocca, J. F., Gillett, N. P., et al. (2019). The Canadian Earth System Model version 5 (CanESM5.0.3). *Geoscientific Model Development*, 12(11), 4823–4873. <https://doi.org/10.5194/gmd-12-4823-2019>
- 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](https://doi.org/10.1016/S0967-0645(02)00003-6)
- Takahashi, T., Sutherland, S. C., Wanninkhof, R., Sweeney, C., Feely, R. A., Chipman, D. W., et al. (2009). Climatological mean and decadal change in surface ocean pCO₂, and net sea-air CO₂ flux over the global oceans. *Deep-Sea Research Part II: Topical Studies in Oceanography*, 56(8–10), 554–577. <https://doi.org/10.1016/j.dsr2.2008.12.009>
- Takeshita, Y., Johnson, K. S., Martz, T. R., Plant, J. N., & Sarmiento, J. L. (2018). Assessment of autonomous pH measurements for determining surface seawater partial pressure of CO₂. *Journal of Geophysical Research: Oceans*, 123(6), 4003–4013. <https://doi.org/10.1029/2017JC013387>
- 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>
- Terhaar, J., Fr  licher, T. L., & Joos, F. (2021). Southern Ocean anthropogenic carbon sink constrained by sea surface salinity. *Science Advances*, 7(18), eabd5964. <https://doi.org/10.1126/sciadv.abd5964>
- Terhaar, J., Goris, N., M  ller, J. D., DeVries, T., Gruber, N., Hauck, J., et al. (2024). Assessment of global ocean biogeochemistry models for ocean carbon sink estimates in RECCAP2 and recommendations for future studies. *Journal of Advances in Modeling Earth Systems*, 16(3), e2023MS003840. <https://doi.org/10.1029/2023MS003840>
- Tjiputra, J. F., Roelandt, C., Bentsen, M., Lawrence, D. M., Lorentzen, T., Schwinger, J., et al. (2013). Evaluation of the carbon cycle components in the Norwegian Earth System Model (NorESM). *Geoscientific Model Development*, 6(2), 301–325. <https://doi.org/10.5194/gmd-6-301-2013>
- van Heuven, S., Pierrot, D., Rae, J., Lewis, E., & Wallace, D. (2011). CO₂SYN v 1.1, MATLAB program developed for CO₂ system calculations. ORNL/CDIAC-105b. Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, U.S. DoE. Retrieved from <https://www.ncei.noaa.gov/access/ocean-carbon-acidification-data-system/oceans/CO2SYS/co2rprt.html>
- Verdy, A., & Mazloff, M. R. (2017). A data assimilating model for estimating Southern Ocean biogeochemistry. *Journal of Geophysical Research: Oceans*, 122(9), 1–22. <https://doi.org/10.1002/2016JC012650>
- Vichi, M., Pinardi, N., & Masina, S. (2007). A generalized model of pelagic biogeochemistry for the global ocean ecosystem. Part I: Theory. *Journal of Marine Systems*, 64(1–4), 89–109. <https://doi.org/10.1016/j.jmarsys.2006.03.006>
- Volk, T., & Hoffert, M. I. (1985). Ocean carbon pumps: Analysis of relative strengths and efficiencies in ocean-driven atmospheric CO₂ changes. In E. Sundquist & W. Broecker (Eds.), *Geophysical Monograph Series* (pp. 99–110). American Geophysical Union. <https://doi.org/10.1029/GM032p0099>
- Volodin, E. M., Mortikov, E. V., Kostykin, S. V., Galin, V. Y., Lykossov, V. N., Gritsun, A. S., et al. (2018). Simulation of the modern climate using the INM-CM48 climate model. *Russian Journal of Numerical Analysis and Mathematical Modelling*, 33(6), 367–374. <https://doi.org/10.1515/rnam-2018-0032>
- 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>
- Watanabe, S., Hajima, T., Sudo, K., Nagashima, T., Takemura, T., Okajima, H., et al. (2011). MIROC-ESM 2010: Model description and basic results of CMIP5-20c3m experiments. *Geoscientific Model Development*, 4(4), 845–872. <https://doi.org/10.5194/gmd-4-845-2011>
- Westberry, T., Behrenfeld, M. J., Siegel, D. A., & Boss, E. (2008). Carbon-based primary productivity modeling with vertically resolved photoacclimation. *Global Biogeochemical Cycles*, 22(2), GB2024. <https://doi.org/10.1029/2007GB003078>
- Williams, N. L., Juranek, L., Feely, R., Russell, J., Johnson, K., & Hales, B. (2018). Assessment of the carbonate chemistry seasonal cycles in the Southern Ocean from persistent observational platforms. *Journal of Geophysical Research: Oceans*, 123(7), 4833–4852. <https://doi.org/10.1029/2017JC012917>
- Williams, N. L., Juranek, L. W., Feely, R. A., Johnson, K. S., Sarmiento, J. L., Talley, L. D., et al. (2017). Calculating surface ocean pCO₂ from biogeochemical Argo floats equipped with pH: An uncertainty analysis. *Global Biogeochemical Cycles*, 31(3), 591–604. <https://doi.org/10.1002/2016GB005541>

- Wu, Y., & Qi, D. (2023). The controversial Southern Ocean air-sea CO₂ flux in the era of autonomous ocean observations. *Science Bulletin*, 68(21), 2519–2522. <https://doi.org/10.1016/j.scib.2023.08.059>
- Yukimoto, S., Kawai, H., Koshiro, T., Oshima, N., Yoshida, K., Urakawa, S., et al. (2019). The Meteorological Research Institute Earth System Model version 2.0, MRI-ESM2.0: Description and basic evaluation of the physical component. *Journal of the Meteorological Society of Japan. Series II*, 97(5), 931–965. <https://doi.org/10.2151/jmsj.2019-051>
- Zahariev, K., Christian, J. R., & Denman, K. L. (2008). Preindustrial, historical, and fertilization simulations using a global ocean carbon model with new parameterizations of iron limitation, calcification, and N₂ fixation. *Progress in Oceanography*, 77(1), 56–82. <https://doi.org/10.1016/j.pocean.2008.01.007>