JAMES | Journal of Advances in Modeling Earth Systems

RESEARCH ARTICLE
10.1029/2023MS003840

Assessment of Global Ocean Biogeochemistry Models for Ocean Carbon Sink Estimates in RECCAP2 and Recommendations for Future Studies

Jens Terhaar[1,2,3], Nadine Goris[4], Jens D. Müller[5], Tim DeVries[6,7], Nicolas Gruber[5], Judith Hauck[8], Fiz F. Perez[9], and Roland Séférian[10]

[1]Department of Marine Chemistry and Geochemistry, Woods Hole Oceanographic Institution, Woods Hole, MA, USA, [2]Climate and Environmental Physics, Physics Institute, University of Bern, Bern, Switzerland, [3]Oeschger Centre for Climate Change Research, University of Bern, Bern, Switzerland, [4]NORCE Climate & Environment, Bjerknes Centre for Climate Research, Bergen, Norway, [5]Environmental Physics, Institute of Biogeochemistry and Pollutant Dynamics, ETH Zurich, Zürich, Switzerland, [6]Department of Geography, University of California, Santa Barbara, Santa Barbara, CA, USA, [7]Earth Research Institute, University of California, Santa Barbara, Santa Barbara, CA, USA, [8]Alfred-Wegener-Institut, Helmholtz-Zentrum für Polar- und Meeresforschung, Bremerhaven, Germany, [9]Instituto de Investigaciones Marinas (IIM), CSIC, Vigo, Spain, [10]CNRM (Université de Toulouse, Météo-France, CNRS), Toulouse, France

Abstract The ocean is a major carbon sink and takes up 25%–30% of the anthropogenically emitted CO_2. A state-of-the-art method to quantify this sink are global ocean biogeochemistry models (GOBMs), but their simulated CO_2 uptake differs between models and is systematically lower than estimates based on statistical methods using surface ocean pCO_2 and interior ocean measurements. Here, we provide an in-depth evaluation of ocean carbon sink estimates from 1980 to 2018 from a GOBM ensemble. As sources of inter-model differences and ensemble-mean biases our study identifies (a) the model setup, such as the length of the spin-up, the starting date of the simulation, and carbon fluxes from rivers and into sediments, (b) the simulated ocean circulation, such as Atlantic Meridional Overturning Circulation and Southern Ocean mode and intermediate water formation, and (c) the simulated oceanic buffer capacity. Our analysis suggests that a late starting date and biases in the ocean circulation cause a too low anthropogenic CO_2 uptake across the GOBM ensemble. Surface ocean biogeochemistry biases might also cause simulated anthropogenic fluxes to be too low, but the current setup prevents a robust assessment. For simulations of the ocean carbon sink, we recommend in the short-term to (a) start simulations at a common date before the industrialization and the associated atmospheric CO_2 increase, (b) conduct a sufficiently long spin-up such that the GOBMs reach steady-state, and (c) provide key metrics for circulation, biogeochemistry, and the land-ocean interface. In the long-term, we recommend improving the representation of these metrics in the GOBMs.

Plain Language Summary In this study, we evaluate the performance of state-of-art global ocean biogeochemistry models (GOBMs) in simulating CO_2 fluxes across the air-sea interface from 1980 to 2018 for the Global Carbon Budget. Across these GOBMs, the simulated CO_2 uptake is systematically lower than that of observation-based estimates and the estimates differ also substantially between GOBMs. As reasons for the too low carbon sink of the GOBMs, we find that the simulations of several GOBMs were initialized after the start of the industrial revolution and that the majority of the considered GOBMs underestimate the large-scale ocean circulation in the Atlantic. The different initialization times of the simulations as well as different strengths of the simulated ocean circulation across the global ocean also partly explain the inter-model differences for the ocean carbon sink. Our analysis of the influence of GOBM dynamics on their simulated carbon sink was impeded by the fact that not all GOBMs had the same initial stability and that the riverine component of the ocean carbon sink is highly uncertain in both observations and GOBMs. Based on our evaluation, we give recommendations for follow up studies.

Nadine Goris and Jens Terhaar should be considered joint first authors.

Key Points:

- The simulated CO_2-uptake by global ocean biogeochemistry models (GOBMs) in the second phase of the REgional Carbon Cycle Assessment and Processes project is systematically lower than observation-based estimates
- This underestimation can, to first order, be explained by the simulation setup as well as biases in surface chemistry and ocean circulation
- Concrete steps forward are proposed to improve simulations of the ocean carbon sink by GOBMs

Supporting Information:

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

Correspondence to:

J. Terhaar,
jens.terhaar@unibe.ch

Citation:

Terhaar, J., Goris, N., Müller, 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*, e2023MS003840. https://doi.org/10.1029/2023MS003840

Received 30 MAY 2023
Accepted 2 FEB 2024

Author Contributions:

Conceptualization: Jens Terhaar, Nadine Goris
Data curation: Jens Terhaar, Nadine Goris, Jens D. Müller
Formal analysis: Jens Terhaar, Nadine Goris, Jens D. Müller

1. Introduction

Currently, the global ocean takes up 25%–30% of all human-made CO_2 emissions (DeVries, 2014; Friedlingstein et al., 2022; Gruber, Clement, et al., 2019; Gruber et al., 2023; Khatiwala et al., 2009; Müller et al., 2023; Terhaar et al., 2022), thereby reducing the growth of atmospheric CO_2 and slowing down global warming (IPCC, 2021).

Investigation: Jens Terhaar, Nadine Goris, Jens D. Müller
Methodology: Jens Terhaar, Nadine Goris, Jens D. Müller, Nicolas Gruber, Judith Hauck, Roland Séférian
Project administration: Jens D. Müller, Nicolas Gruber, Judith Hauck
Visualization: Jens Terhaar, Nadine Goris, Jens D. Müller
Writing – original draft: Jens Terhaar, Nadine Goris
Writing – review & editing: Jens Terhaar, Nadine Goris, Jens D. Müller, Tim DeVries, Nicolas Gruber, Judith Hauck, Fiz F. Perez, Roland Séférian

However, the additional carbon in the ocean causes ocean acidification (Haugan & Drange, 1996) and reduces the efficiency of the ocean carbon sink (Broecker et al., 1979; Revelle & Suess, 1957).

The main driver of the evolution of the global ocean carbon sink from preindustrial times to present is the increasing atmospheric CO_2 due to human activity (Sarmiento et al., 1992). The additional dissolved inorganic carbon (DIC) in the ocean due to rising atmospheric CO_2 concentrations is known as anthropogenic carbon (C_{ant}; Sarmiento et al., 1992), while the DIC that existed prior to the start of the industrial revolution is called natural carbon (C_{nat}). Second order terms in the historical evolution of the ocean carbon sink are climate-change and climate-variability driven changes in the anthropogenic and natural air-sea CO_2 fluxes (Joos et al., 1999; Le Quéré et al., 2000; McNeil & Matear, 2013), as well as anthropogenic changes in the riverine carbon fluxes (Regnier et al., 2013; Terhaar et al., 2022). At the global scale, the air-sea C_{ant} flux is controlled by the rate of C_{ant} transport from the surface ocean to the deep ocean, which depends on the concentration of C_{ant} in the surface ocean (Broecker et al., 1979) and the surface-to-deep water volume transport (Caldeira & Duffy, 2000; Mikaloff Fletcher et al., 2006; Orr et al., 2001; Sarmiento et al., 1992). In contrast, the air-sea flux of C_{nat} is primarily controlled by the interaction of surface heating or cooling affecting the solubility of CO_2 in seawater and transport and mixing, and biological processes of photosynthesis, respiration, and $CaCO_3$ production (Sarmiento & Gruber, 2006). While there is agreement on these drivers for C_{ant} and C_{nat} fluxes and their relative importance, an accurate quantification of the carbon sink and its processes is still challenging.

More than 100 scientists around the globe have worked on providing an updated quantification of the carbon fluxes between the atmosphere, land, and ocean during Phase 2 of the REgional Carbon Cycle Assessment and Processes project (RECCAP2) (Poulter et al., 2022). The ocean part of RECCAP2 assesses the most up-to-date air-sea carbon flux estimates based on statistical methods applied to observations of surface ocean partial pressure of CO_2 (pCO_2 products) and hindcast simulations from global and regional ocean biogeochemistry models (ROCMs and GOBMs) to better understand the global and regional ocean carbon sink over the last three decades, its decadal and inter-annual variability and seasonal cycle, and the contribution of the biological pump. Although they contain data from similar GOBMs and pCO_2 products, the compiled database of RECCAP2 (Müller, 2023) goes well beyond that used in the framework of the Global Carbon Budget (Friedlingstein et al., 2022). Specifically, the RECCAP2 database contains simulation results from a broader set of numerical simulations, and it includes much more spatially and temporally refined data and many more variables. This database permits us to analyze the spatially and temporally resolved air-sea CO_2 fluxes and the processes controlling the ocean carbon sink. With this study here, we provide an evaluation of the GOBM hindcast simulations to better contextualize the model results in the different studies of the AGU special issue "REgional Carbon Cycle Assessment and Processes—2 (RECCAP2)" and in the 2020 and 2022 edition of the Global Carbon Budget (Friedlingstein et al., 2020, 2022) and to make recommendations for future assessments of the ocean carbon sink using GOBMs.

The RECCAP2 project is a continuation of the large efforts that have been undertaken in the last decades to quantify the past and present ocean carbon sink with pCO_2 products (Chau et al., 2022; Gregor & Gruber, 2021; Gregor et al., 2019; Iida et al., 2021; Landschützer et al., 2014; Rödenbeck et al., 2013; Watson et al., 2020; Zeng et al., 2014) and GOBMs forced with historic atmospheric reanalysis data (Hauck et al., 2020; Orr et al., 2001; Sarmiento et al., 1992; Sarmiento & Sundquist, 1992). The global ocean carbon sink estimates differ across the different methods and models. In the multi-model mean, the simulated change in the air-sea CO_2 flux since preindustrial times as reported by the Global Carbon Budget is consistently less negative (lower uptake) than the mean estimate of the pCO_2-products (1990s: -1.91 ± 0.25 Pg C year^{-1} in models vs. -2.14 ± 0.34 Pg C year^{-1} for pCO_2 products, 2000s: -2.05 ± 0.27 Pg C year^{-1} vs. -2.34 ± 0.21 Pg C year^{-1}, and 2010s: -2.42 ± 0.29 Pg C year^{-1} vs. -3.02 ± 0.22 Pg C year^{-1}; Friedlingstein et al., 2022). This difference between the models and pCO_2 products in the 2010s is around half as large as the annual CO_2 emissions in the United States of America over the same period (Friedlingstein et al., 2022). This highlights the need for a more rigorous quantification of the ocean carbon sink to fully close the global carbon budget (Hauck et al., 2020). A better understanding of the fidelity of GOBMs is also needed if such models are to be used for monitoring, reporting, and verification of ocean-based carbon dioxide removal techniques (Gattuso et al., 2018).

Prior GOBM intercomparison studies (Khatiwala et al., 2013; Orr et al., 2001; Wanninkhof et al., 2013) and studies with related Earth System Models (ESMs) suggest several reasons for the differences mentioned above. Among them are biases in model dynamics such as the mode, intermediate, and deep-water formation in the North Atlantic (Goris et al., 2018, 2023; Terhaar et al., 2022) and Southern Ocean (Bourgeois et al., 2022; Fu

et al., 2022; Terhaar, Frölicher, & Joos, 2021; Terhaar et al., 2022), both causing a bias in the amount of carbon that is transported from the surface to the deep ocean. Also biases in the model ocean carbonate chemistry affect the anthropogenic CO_2 uptake (Terhaar et al., 2022; Vaittinada Ayar et al., 2022). Other reasons for the above-mentioned differences are related to the setup of the model simulations. For example, the starting date of model simulations is often several decades delayed relative to the onset of the anthropogenic CO_2 increase in the atmosphere due to changes in land-use and the start of the industrial revolution in the second half of the eighteenth century (Bronselaer et al., 2017; Terhaar, Orr, Gehlen, et al., 2019), leading to a too low ocean carbon uptake and storage. Associated with the setup of model simulations is also the spin-up procedure (Séférian et al., 2016), where a too short spin-up can lead to model drift and adds a significant source of uncertainty to the multi-model spread. Based on these findings, the here presented study identifies inter-model differences between GOBM simulations of the natural and anthropogenic components of the ocean carbon sink as well as differences between the ocean carbon sink estimates of GOBMs and pCO_2 products at a global and regional level. We also investigate the underlying reasons for these differences and provide recommendations for future assessments of the ocean carbon sink using GOBMs.

2. Materials and Methods

2.1. Ocean Biogeochemistry Models

The GOBMs analyzed in this study are general ocean circulation models with coupled sea ice and ocean biogeochemistry model components. They simulate the transport of biogeochemical tracers through advection and mixing and their cycling through biogeochemical processes (primary production, grazing, remineralization, etc., see Fennel et al., 2022). The air-sea CO_2 flux in these models is based on the simulated ocean carbon dynamics and the prescribed atmospheric CO_2 mixing ratio. In this study, we analyzed the following eight GOBMs in full: CESM-ETHZ (Yang & Gruber, 2016), CNRM-ESM2-1 (Séférian et al., 2019), EC-Earth3 (Döscher et al., 2022), FESOM- RecoM-LR (Hauck et al., 2020), MRI-ESM2-1 (Urakawa et al., 2020), NorESM-OC1.2 (Schwinger et al., 2016), ORCA025-GEOMAR (Physics are described in Madec et al. (2017), and biogeochemistry in Chien et al. (2022)) and ORCA1-LIM3-PISCES (Aumont et al., 2015). Three GOBMs that submitted data to RECCAP2 were not or only partially included in our analyses: The MPI-OM-HAMOCC model (Mauritsen et al., 2019) was not used here as the separation into all individual flux components (see Section 2.2.3) was not possible because its different simulations were forced with different atmospheric forcing data sets. Similarly, MOM6-Princeton (Stock et al., 2020) did not perform two of the RECCAP2-simulations, preventing us from diagnosing the individual CO_2 flux components. Therefore, we do not consider MOM6-Princeton when presenting values or plots for the GOBM-ensemble to conserve consistency between the different flux components. But we present its results separately when possible. Lastly, PlankTOM12 (Wright et al., 2021) has strong salinity biases in all basins. These biases and associated biases in circulation lead to an anthropogenic carbon storage pattern that does not resemble any of the observation-based estimates. While we plot its results in the supplementary Figures of Section 3.3.1 (Interior Ocean anthropogenic carbon storage) and explain the reasons for its exclusion there, we exclude it from all GOBM results in terms of multi-model mean and standard deviation. The exclusion of these three GOBMs in GOBM-ensemble values can lead to this study presenting slightly different regional estimates than other RECCAP2-studies (DeVries et al., 2023; Hauck, Gregor, et al., 2023; Hauck, Nissen, et al., 2023; Pérez et al., 2023; Yasunaka et al., 2023).

The here-considered GOBMs were forced with atmospheric fields, such as atmospheric temperature and wind speeds, from different versions of either the Japanese Reanalysis JRA-55-do (Tsujino et al., 2018) or of the reanalysis from NCEP/NCAR (Large & Yeager, 2009). Details of the respective model resolutions, forcings, and references are listed in Supplementary Table 1 in DeVries et al. (2023). As our study analyzed the influence of the simulated Atlantic Meridional Overturning Circulation (AMOC) on the simulated sea-air carbon fluxes, we additionally considered a second realization of the RECCAP2-simulations by the model CESM-ETHZ with a different sea surface salinity restoring. In the standard realization of the CESM-ETHZ simulations, the salinity restoring timescale was 2 years everywhere at the ocean surface, whereas the second realization used a timescale of 300 days north of 45°S and of 60 days south of 45°S. The shortened restoring timescale in the Southern Ocean better captures seasonal freshwater fluxes due to sea ice formation and melting. This change in the salinity restoring led to an improvement of the overturning circulation simulated by CESM-ETHZ, not only in the Southern Ocean, but also in the North Atlantic. Here, the previously very weak AMOC at 26°N increased from 3.5 to 14.4 Sv (averaged over years 2005–2018).

2.2. Sea-Air CO_2 Flux

2.2.1. Sign Convention

Throughout this study, the CO_2 flux between the atmosphere and ocean is defined as a sea-to-air flux, thus with a negative flux indicating an uptake of CO_2 by the ocean and a positive flux indicating an outgassing. Positive land-to-sea riverine fluxes indicate a flux into the ocean and positive sea-to-sediment burial fluxes indicate a flux from the ocean into the sediments.

2.2.2. Components of the Sea-Air CO_2 Flux

In this study, we aim at quantifying the net sea-air CO_2 flux (F^{net}), that is, the contemporary air-sea CO_2 flux and not the anthropogenic-perturbation of the pre-industrial air-sea CO_2 fluxes as quantified in the Global Carbon Budget (Friedlingstein et al., 2022). F^{net} can be divided into five components. The first two components are part of the anthropogenic sea-air CO_2 flux from increasing atmospheric CO_2 in the atmosphere (F_{ant}), which was further divided into a steady-state component $F_{\text{ant}}{}^{\text{ss}}$ representing the anthropogenic sea-air CO_2 flux in the absence of climate change and climate variability, and into a non-steady state component $F_{\text{ant}}{}^{\text{ns}}$ reflecting the effect of climate change and climate variability on F_{ant}. The next two components are part of the natural sea-air flux of CO_2 under pre-industrial atmospheric CO_2 (F_{nat}), which was also divided into F_{nat} under a constant climate (steady-state F_{nat} or short $F_{\text{nat}}{}^{\text{ss}}$), and the modulation of the F_{nat} due to climate variability and climate change (non-steady state F_{nat} or short $F_{\text{nat}}{}^{\text{ns}}$). The fifth flux component is the sea-air CO_2 flux due to the carbon, alkalinity, and nutrient fluxes from rivers and coastal erosion and their respective burial in sediments ($F_{\text{nat}}{}^{\text{riv-bur}}$) (Lacroix et al., 2020). While previous literature has often called this a riverine-induced flux, we decided to call it riverine-burial induced flux to emphasize that the flux depends on both, the carbon flux from rivers into the ocean and the carbon flux into the sediments. Some of the other papers of the AGU special issue "Regional Carbon Cycle Assessment and Processes —2 (RECCAP2)" consider $F_{\text{nat}}{}^{\text{riv-bur}}$ to be an integral part of $F_{\text{nat}}{}^{\text{ss}}$. We kept them separated to the degree that this is possible in order to analyze all flux components individually.

The net flux across the sea-air interface (F^{net}) can thus be written as:

$$F^{\text{net}} = F_{\text{ant}}{}^{\text{ss}} + F_{\text{ant}}{}^{\text{ns}} + F_{\text{nat}}{}^{\text{ss}} + F_{\text{nat}}{}^{\text{ns}} + F_{\text{nat}}{}^{\text{riv-bur}}. \quad (1)$$

Throughout this paper, carbon inventories are referred to as "C" in analogy to the fluxes that are abbreviated with "F." The same indices as for the fluxes were used to distinguish the respective components of carbon inventories and their change over time. While RECCAP2 mainly aimed at quantifying F^{net}, the annually released Global Carbon Budget aims at quantifying the anthropogenic perturbation of the preindustrial air-sea CO_2 flux. This change in F^{net} is called S^{OCEAN} in the Global Carbon Budget.

2.2.3. RECCAP2 Simulations and Their Relation to CO_2 Flux Components

The RECCAP2 database (Müller, 2023) provides model output from 1980 to 2018 from four simulations (called simulations A, B, C and D) that aim to quantify the different components of the oceanic CO_2 flux.

- Simulation A is forced with historical atmospheric reanalysis data and historically increasing CO_2, yielding:

$$F^{\text{SimA}} = F_{\text{ant}}{}^{\text{ss}} + F_{\text{ant}}{}^{\text{ns}} + F_{\text{nat}}{}^{\text{ss}} + F_{\text{nat}}{}^{\text{ns}} + F_{\text{nat}}{}^{\text{riv-bur}} + F^{\text{drift+bias}}, \quad (2)$$

  where $F^{\text{drift+bias}}$ is the bias in the simulated sea-air CO_2 flux due to insufficient spin up and its drift over time (more details on this can be found at the end of this section).
- Simulation B is forced with the same repeated annual atmospheric forcing and constant pre-industrial CO_2 levels, yielding:

$$F^{\text{SimB}} = F_{\text{nat}}{}^{\text{ss}} + F_{\text{nat}}{}^{\text{riv-bur}} + F^{\text{drift+bias}}. \quad (3)$$

- Simulation C is forced with a constant atmospheric forcing and historically increasing CO_2, yielding:

$$F^{\mathrm{SimC}} = F_{\mathrm{ant}}{}^{\mathrm{ss}} + F_{\mathrm{nat}}{}^{\mathrm{ss}} + F_{\mathrm{nat}}{}^{\mathrm{riv\text{-}bur}} + F^{\mathrm{drift+bias}}. \quad (4)$$

- Simulation D is forced with historical atmospheric reanalysis data and constant pre-industrial CO_2 levels, yielding:

$$F^{\mathrm{SimD}} = F_{\mathrm{nat}}{}^{\mathrm{ss}} + F_{\mathrm{nat}}{}^{\mathrm{ns}} + F_{\mathrm{nat}}{}^{\mathrm{riv\text{-}bur}} + F^{\mathrm{drift+bias}}. \quad (5)$$

Simulations with a constant atmospheric climate (B, C) represent steady-state processes, while simulations with a variable climate (A, D) represent both steady-state and non-steady state processes. Similarly, simulations with rising CO_2 (A, C) represent both natural and anthropogenic CO_2 fluxes, while simulations with pre-industrial CO_2 (B, D) represent only natural CO_2 fluxes. For each GOBM, fluxes across the land-sea and sediment interface are the same in simulations A, B, C, and D, so that we assume $F_{\mathrm{nat}}{}^{\mathrm{riv\text{-}bur}}$ also to be the same in each simulation. Across the here considered GOBMs, two GOBMs have no riverine input at all whereas the remaining GOBMs prescribe monthly climatological riverine input of carbon (see Section 3.1.1).

Although RECCAP2 focuses on the time from 1980 to 2018, the four simulations all start in preindustrial times. However, different definitions of "preindustrial" were used such that simulation start years varied between 1765 and 1870, and the corresponding assumed pre-industrial CO_2 mixing ratios between 278 and 286 ppm, depending on the associated GOBM. Furthermore, all four simulations were forced with a repeated (normal year) atmospheric forcing until historical atmospheric reanalysis fields became available in 1948 or 1958 (depending on the atmospheric reanalysis that was used to force the GOBM).

Four of the 10 GOBMs considered here (FESOM-REcoM-LR, MOM6-Princeton, ORCA1-LIM3-PISCES, PlankTOM12) ran the four simulations without a pre-industrial spin-up. The remaining six GOBMs (CESM-ETHZ, CNRM-ESM2-1, EC-Earth3, MRI-ESM2-1, NorESM-OC1.2, and ORCA025-GEOMAR) performed a pre-industrial spin-up that lasted between 137 and 1,586 years (overview in Supplementary Table 1 in DeVries et al. (2023)) with the goal to reach a steady-state where all multi-annual mean fluxes are time-invariant at the local scale and globally integrated zero. Few of the six models with spin-up reach this objective, largely because of the spin-up being too short compared to the century time-scale of global ocean overturning. This too short spin-up (or the complete lack thereof) leads to a model not reaching steady-state and can cause a substantial spin-up related bias in the simulated air-sea CO_2 fluxes (Griffies et al., 2016; Orr et al., 2017; Séférian et al., 2016). The models analyzed here have global CO_2 flux biases ranging between −0.35 and 0.17 Pg C year^{-1}, with a relatively small drift over time (Hauck et al., 2020). However, regionally, this effect can be more important. This $F^{\mathrm{drift+bias}}$ does not include other biases in the sea-air CO_2-flux stemming from errors in ocean circulation or biogeochemistry.

2.2.4. Estimating the Sea-Air CO_2 Flux and Its Components From RECCAP2 Simulations

Three components of the sea-air CO_2 flux can be estimated globally and regionally by subtracting the sea-air CO_2 flux in one RECCAP2 simulation from the sea-air CO_2 flux in another RECCAP2 simulation, assuming that $F_{\mathrm{nat}}{}^{\mathrm{riv\text{-}bur}}$ and $F^{\mathrm{drift+bias}}$ are not affected by increasing atmospheric CO_2 or changing atmospheric forcing across the varying simulations and that the different flux components add up without substantial non-linearities:

$$F_{\mathrm{ant}}{}^{\mathrm{ss}} = F^{\mathrm{SimC}} - F^{\mathrm{SimB}}, \quad (6)$$

$$F_{\mathrm{ant}}{}^{\mathrm{ns}} = F^{\mathrm{SimA}} - F^{\mathrm{SimC}} + F^{\mathrm{SimB}} - F^{\mathrm{SimD}}, \quad (7)$$

$$F_{\mathrm{nat}}{}^{\mathrm{ns}} = F^{\mathrm{SimD}} - F^{\mathrm{SimB}}. \quad (8)$$

Within the setup of RECCAP2 and the here considered GOBMs, there is no combination of simulations that allows to isolate $F_{\mathrm{nat}}{}^{\mathrm{ss}}$. When inserting Equations 6–8 into Equation 1, the global net air-sea CO_2 flux ($F_{\mathrm{global}}{}^{\mathrm{net}}$) can be estimated as follows:

$$F_{\text{global}}^{\text{net}} = F^{\text{SimA}} - F^{\text{SimB}} + F_{\text{nat}}^{\text{ss}} + F_{\text{nat}}^{\text{riv-bur}}. \quad (9)$$

As $F_{\text{nat}}^{\text{ss}}$ is globally by definition zero, only $F_{\text{nat}}^{\text{riv-bur}}$ has to be known for a GOBM-based estimate of $F_{\text{global}}^{\text{net}}$. Unfortunately, $F_{\text{nat}}^{\text{riv-bur}}$ cannot be quantified from the here-used GOBM simulations because these have an inadequate (seven GOBMs) or no (two GOBMs) representation of carbon riverine input and carbon sediment burial resulting in small to inexistent $F_{\text{nat}}^{\text{riv-bur}}$ (see Section 3.1.1.). To allow for a reasonable comparison between observation-based and GOBM-based estimates of $F_{\text{global}}^{\text{net}}$, we hence replace the simulated $F_{\text{nat}}^{\text{riv-bur}}$ with an observation-based estimate $F_{\text{obs}}^{\text{riv-bur}}$ of 0.65 ± 0.15 Pg C year^{-1} (based on Regnier et al. (2022) but we utilize the 1-sigma uncertainty instead of the 2-sigma uncertainty). For comparisons between observation-based and modeled estimates of air-sea CO_2 fluxes, river adjustments have been frequently used. For example, in the Global Carbon Budget (Friedlingstein et al., 2022), $F_{\text{obs}}^{\text{riv-bur}}$ is subtracted from the pCO_2 products to estimate S_{OCEAN}. As we do not estimate S_{OCEAN} but F^{net}, we do subtract $F_{\text{obs}}^{\text{riv-bur}}$ from the pCO_2 products but also use $F_{\text{obs}}^{\text{riv-bur}}$ as an adjustment for the global GOBM-estimate. As $F_{\text{nat}}^{\text{riv-bur}}$ is the same across all simulations in each GOBM, this approximation affects solely the comparison of F^{net} between GOBMs and pCO_2 products and not the model evaluation of each individual flux component (Equations 6–8). However, this approximation disregards that land-sea riverine and burial fluxes change over time (Regnier et al., 2013; Séférian et al., 2019; Terhaar et al., 2022) and that these changes affect the sea-air CO_2 flux regionally (Gomez et al., 2021; Terhaar, Orr, Ethé, et al., 2019) and globally (Regnier et al., 2013; Terhaar et al., 2022). As there is no observation-based estimate of the temporally resolved riverine-burial-induced fluxes yet, we cannot quantify the effect of this approximation.

Regionally, estimating F^{net} from the RECCAP2 simulations is more difficult, because the regional $F_{\text{nat}}^{\text{ss}}$ is not zero as the ocean takes up and releases natural carbon regionally. Therefore, regional F^{net} ($F_{\text{regional}}^{\text{net}}$) cannot be estimated with Equation 9. Hence, we estimate $F_{\text{regional}}^{\text{net}}$ from simulation A plus an added observation-based estimate of the regional $F_{\text{nat}}^{\text{riv-bur}}$ ($F_{\text{obs}}^{\text{riv-bur}}$) and accept the simulated regional $F_{\text{nat}}^{\text{drift+bias}}$ and $F_{\text{nat}}^{\text{riv-bur}}$ as inherent uncertainties:

$$F_{\text{regional}}^{\text{net}} = F^{\text{SimA}} + F_{\text{obs}}^{\text{riv-bur}} = F_{\text{ant}}^{\text{ss}} + F_{\text{ant}}^{\text{ns}} + F_{\text{nat}}^{\text{ss}} + F_{\text{nat}}^{\text{ns}} + F_{\text{nat}}^{\text{riv-bur}} + F^{\text{drift+bias}} + F_{\text{obs}}^{\text{riv-bur}} \quad (10)$$

This $F_{\text{regional}}^{\text{net}}$ estimate hence includes both $F_{\text{nat}}^{\text{riv-bur}}$ from the GOBMs as well as the observational estimate $F_{\text{obs}}^{\text{riv-bur}}$, based on the assumption that the $F_{\text{nat}}^{\text{riv-bur}}$ from the GOBMs is inadequately low. According to RECCAP2 glossary, the regional observation-based estimate of $F_{\text{obs}}^{\text{riv-bur}}$ should be derived from the estimated regional pattern of $F_{\text{nat}}^{\text{riv-bur}}$ (Lacroix et al., 2020), which is scaled with a constant factor (2.83) for all grid cells such that the global integral matches the postulated global value of $F_{\text{obs}}^{\text{riv-bur}}$ of 0.65 Pg C year^{-1}. Overall, the impossibility to disentangle the regional values of $F_{\text{nat}}^{\text{ss}}$, $F^{\text{drift+bias}}$, and $F_{\text{nat}}^{\text{riv-bur}}$ in the models and the uncertainties of the regional observation-based $F_{\text{obs}}^{\text{riv-bur}}$ hence add additional uncertainty to the regional estimates of F^{net}.

2.3. Observation-Based Estimates, Their Uncertainties, and Their Usage for Comparison With GOBMs

To compare the net sea-air CO_2 fluxes from the GOBMs with observation-based estimates, we utilize the RECCAP2 data set of pCO_2 products (Müller, 2023), including AOML_EXTRAT, CMEMS-LSCE-FFNN, CSIR-ML6, JenaMLS, JMA-MLR, MPI-SOMFFN, OceanSODA-ETHZ, UOEX_Wat20, and NIES-MLR3 (see Supplementary Table 2 in DeVries et al. (2023) for references and further details). These pCO_2 products are a product of statistical models and sparse observations of surface ocean partial pressure of CO_2. We calculate long-term averages and trends over these products for the period 1985 through 2018 for all products except AOM-L_EXTRAT, who did not provide estimates before 1997, and UOEX_Wat20, who did not provide estimates before 1989 and uses a conceptually different surface ocean pCO_2 data set. In addition, the simulated regional $F_{\text{nat}}^{\text{ss}}$ was compared to ocean inversion-based estimates (Mikaloff Fletcher et al., 2007).

As observation-based estimates for the C_{ant}-storage, we used three mapped products which we name after the method that was used to produce them. Hence, we refer to the mapped anthropogenic carbon storage between the years 1800 and 2002 from the GLODAPv2.2016b-product (Lauvset et al., 2016) as TTD-estimate; the mapped $C_{\text{ant}}^{\text{ns+ss}}$-storage from the year 1800–1994 (Sabine et al., 2004) as ΔC^*-estimate and the mapped $C_{\text{ant}}^{\text{ns+ss}}$-storage between years 1994 and 2007 (Gruber, Clement, et al., 2019) as eMLR(C^*)-estimate, respectively. Uncertainties of the globally integrated ΔC^*- and eMLR(C^*)-estimates were provided with the respective estimates. For the

TTD-estimate, a comprehensive error estimate is lacking, and we utilized the error-estimate of ±29% for the C_{ant}-storage of the North Atlantic (Steinfeldt et al., 2009), which is a rather conservative error estimate (Khatiwala et al., 2013; Terhaar, Tanhua, et al., 2020). When referring to a comparison between TTD-, ΔC^*- and eMLR(C^*)-estimates and GOBM-estimates of interior ocean C_{ant}-storage, only regions where all estimates provide data were included, which means that depths under 3,000 m as well as the Arctic Ocean and the marginal Seas were excluded.

The RAPID-Meridional Overturning Circulation and Heatflux Array-Western Boundary Time Series array was used (Frajka-Williams et al., 2021) to calculate an observation-based estimate of the mean AMOC (here defined as maximum of the Atlantic meridional overturning streamfunction at 26°N) strength from 2005 to 2018 as 16.93 Sv. The measurement uncertainty of this value is estimated to be ±0.57 Sv based on the rules of error propagation, where we assume the initial error of the first 10-day measurement to be 1.5 Sv (https://rapid.ac.uk/rapidmoc/rapid_data/README_ERROR.pdf, accessed in October 2022) and each year to be independent as the moorings of the observational array are exchanged every year.

Mean estimates and uncertainties of the interfrontal sea surface salinity and the associated ventilated water volume in the Southern Ocean (as defined in Terhaar, Frölicher, and Joos (2021)) were here calculated for each GOBM and as observation-based estimate using gridded monthly climatologies of salinity and temperature from the World Ocean Atlas 2018 (Locarnini et al., 2018; Zweng et al., 2018).

For comparisons of surface DIC and alkalinity between observation-based estimates and GOBMs, monthly and spatially resolved gridded estimates of DIC and alkalinity provided by OceanSODA-ETHZ (Gregor & Gruber, 2021), CMEMS-LSCE-FFNN (Chau et al., 2022), and JMA-MLR (Iida et al., 2021) based on observations of surface ocean pCO_2 and alkalinity (henceforth called pCO_2/alkalinity products) were used. Furthermore, gridded GLODAPv2 estimates of the same variables were also used (Lauvset et al., 2016), where DIC is normalized to the atmospheric pCO_2 of 2002. For comparison, output from the pCO_2/alkalinity products and GOBMs were averaged over the years 1986–2018, the longest time period available with the year 2002 in its center.

Additionally, we compared the simulated and observation-based estimates of the Revelle factor (Revelle & Suess, 1957), carbonate ion (CO_3^{2-}) concentrations, and the chemical surface ocean uptake capacity. The Revelle factor describes the overall uptake capacity of the ocean:

$$\text{Revelle} = (\partial[pCO_2]/[pCO_2])/(\partial\text{DIC}/\text{DIC}). \quad (11)$$

We re-arranged this equation to quantify the amount of additional carbon that the surface ocean can take up for a given increase in pCO_2 ($\partial\text{DIC}/\partial[pCO_2]$) and defined this to be the chemical uptake capacity:

$$\partial\text{DIC}/\partial[pCO_2] = \text{DIC}/(\text{Revelle} \times [pCO_2]). \quad (12)$$

For consistency, the Revelle factor, CO_3^{2-}, and the chemical uptake capacity were calculated based on the provided temperature, salinity, DIC, and alkalinity in GLODAPv2, the three pCO_2/alkalinity products, and all GOBMs using mocsy2.0 (Orr & Epitalon, 2015) and the equilibrium constants recommended for best practice by Dickson et al. (2007) based on Lueker et al. (2000), Mehrbach et al. (1973), Millero (1995), and Weiss (1974), as well as the borate-salinity relationship from Uppström (1974).

Several of the observation-based estimates described above have been used to constrain the GOBM ensemble within an emergent constraint framework (Boé et al., 2009; Eyring et al., 2019; Hall et al., 2019). To obtain the constrained variables and their uncertainties, we here followed the approach from Cox et al. (2013) that has been frequently used over the recent years in ocean biogeochemistry (Bourgeois et al., 2022; Goris et al., 2018, 2023; Kwiatkowski et al., 2017; Terhaar, Frölicher, & Joos, 2021; Terhaar, Kwiatkowski, & Bopp, 2020; Terhaar et al., 2022; Terhaar, Torres, et al., 2021).

2.4. Uncertainties and Ensemble Spread

We utilized the 1-sigma standard-deviation either across the ensemble of GOBMs or pCO_2 products to describe the uncertainty related to varying methods, modules and parametrizations within the GOBMs or pCO_2 products.

Table 1
Ensemble Mean Estimate of Global and Regional Net Sea-Air CO_2-Fluxes (Pg C Year^{-1}) of Global Ocean Biogeochemistry Models and pCO_2 Products

	Global	Atlantic	Pacific	Indian	Arctic	Southern
GOBMs	−1.41 ± 0.24	−0.23 ± 0.15	−0.34 ± 0.12	−0.10 ± 0.06	−0.06 ± 0.03	−0.73 ± 0.31
pCO_2 products	−1.61 ± 0.21	−0.36 ± 0.06	−0.36 ± 0.11	−0.11 ± 0.02	−0.09 ± 0.05	−0.73 ± 0.06

Note. The GOBM uncertainty excludes the uncertainty of $F_{obs}^{riv\text{-}bur}$.

When globally comparing the simulated F^{net} of the GOBMs to that of the pCO_2 products, $F_{obs}^{riv\text{-}bur}$ was added to the GOBM estimate (see Section 2.2.4). The 1-sigma uncertainty of $F_{obs}^{riv\text{-}bur}$ (±0.15 Pg C year^{-1}) increases the uncertainty of the GOBM-derived estimate. For the global F^{net} estimates from GOBMs, we will therefore provide both a combined uncertainty (standard deviation of GOBM ensemble and of $F_{obs}^{riv\text{-}bur}$) and a pure standard deviation that does not include the uncertainty of $F_{obs}^{riv\text{-}bur}$ and hence is a measure of model-based differences only. Excluding the uncertainty of $F_{obs}^{riv\text{-}bur}$ allows comparing the ensemble spread of estimates of GOBMs to that of the pCO_2 products. Regionally, the uncertainty of F^{net} is only provided as the standard deviation across the GOBM ensemble, because regional uncertainties of $F_{obs}^{riv\text{-}bur}$ are not quantified so far.

2.5. Definition of Ocean Basins and Sub-Basin Biomes

For our analysis, we applied the RECCAP2 biome-mask and the associated definition of ocean basins (Figure S1 in Supporting Information S1). The RECCAP2 biome-mask is a slightly modified version of the oceanic biomes of Fay and McKinley (2014), designed to capture large-scale biogeochemical functioning. In comparison to the original biomes, the RECCAP2 biome mask newly introduces the biomes of the Barents Sea as part of the Arctic and the Mediterranean Sea as part of the Atlantic.

2.6. Quantifying the Underestimation of the Ocean Carbon Sink Due To Different Starting Dates

To quantify the difference in the simulated anthropogenic carbon uptake from 1980 to 2018 due to different starting dates (see Section 3.3.2), it would be ideal to re-run all simulations that started after the anthropogenically caused increase in atmospheric CO_2 from a common pre-industrial starting time. However, spinning-up several GOBMs with another pre-industrial pCO_2 and re-running the historical simulations from an earlier start date is computationally too expensive to be achieved within the framework of RECCAP2. Therefore, we here approximate the magnitude of this underestimation by running two simulations, one starting in 1765 (here assumed to represent the time before the anthropogenically caused increase in atmospheric CO_2 began) and one in 1850 (the year where the GOBM simulations with the latest starting time began). The year 1765 has been chosen as it is defined to be the year where atmospheric CO_2 levels started to increase due to changes in land use (Khatiwala et al., 2009) and as this year has been established in many studies about C_{ant} and F_{ant} (e.g., Khatiwala et al., 2009, 2013; Matsumoto & Gruber, 2005; Mikaloff Fletcher et al., 2006). Moreover, the usage of the year 1765 allows a comparison to a study by Bronselaer et al. (2017).

We perform these simulations with different starting dates with an Earth System Model of Intermediate Complexity (EMIC), Bern3D-LPX (Lienert & Joos, 2018; Roth et al., 2014). The model was used with three different ocean mixing parameters and hence AMOC-strengths to cover the wide range of ocean carbon sink strength across the GOBM ensemble (see Terhaar et al. (2023) for details). We compare this Bern3D-LPX estimate of the underestimation of the carbon uptake due to a later starting date to an estimate of Bronselaer et al. (2017) based on two "offline" approaches: the transport matrix method (Khatiwala et al., 2005) that simulates biogeochemical tracer propagation, and an impulse response function (Joos et al., 2013), which assumes each year's emission as an impulse and quantifies the uptake of ESMs of such an impulse over time. Both approaches consider related changes of the oceanic buffer capacity.

3. Results

For the period 1985 to 2018, the ensemble of eight GOBMs simulates a mean annual globally integrated F^{net} (−1.41 ± 0.24 Pg C year^{-1}; here excluding uncertainties of $F_{obs}^{riv\text{-}bur}$) that is statistically indistinguishable from that estimated by the pCO_2-products (−1.61 ± 0.21 Pg C year^{-1}) (Table 1, Figure 1). In addition, the overall

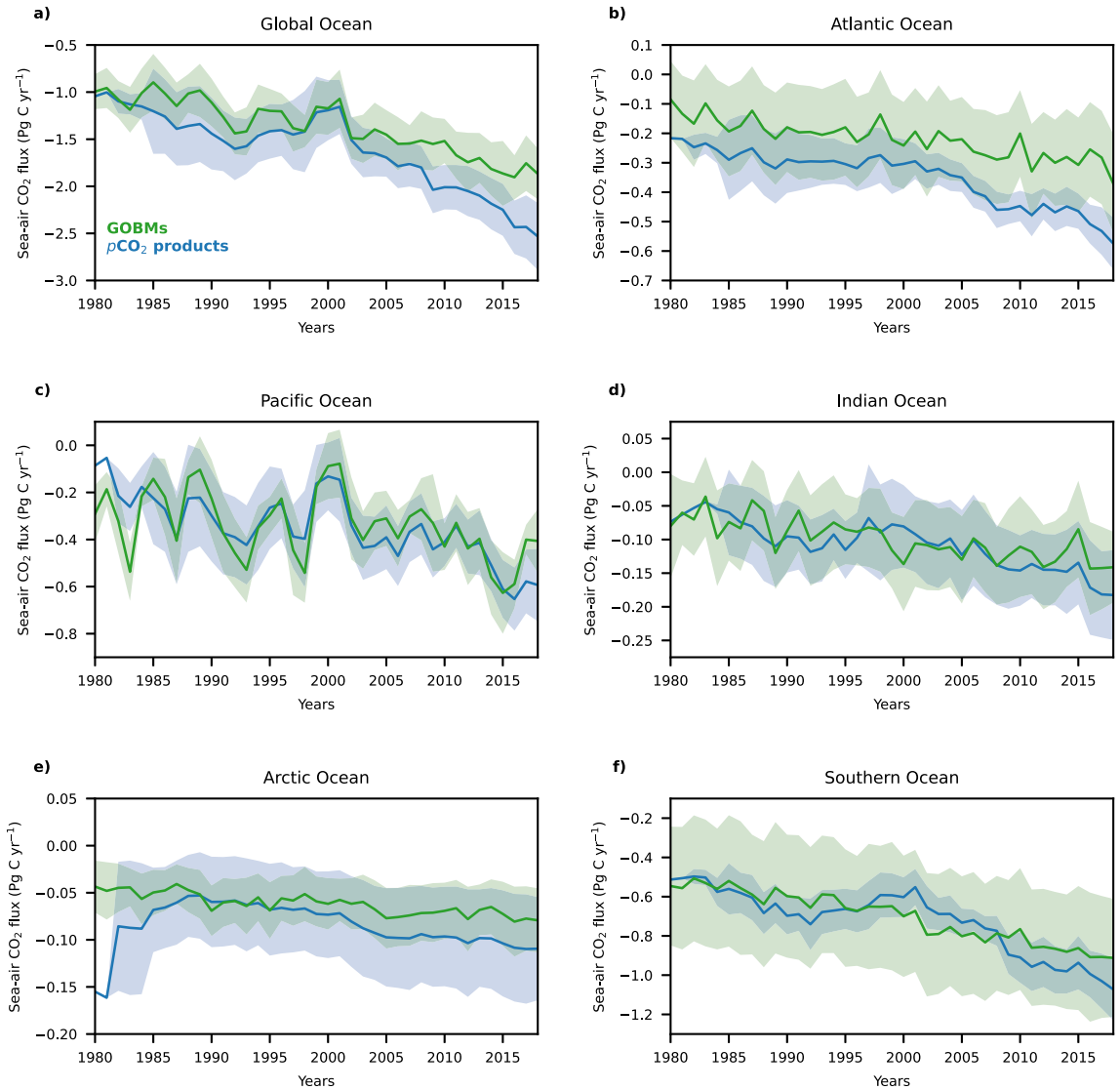

Figure 1. *Time series of global and regional net sea-air CO_2 fluxes from 1980 to 2018 based on global ocean biogeochemistry models (GOBMs) and pCO_2 products*. The average sea-air CO_2 flux from the GOBMs adjusted for the riverine-burial induced sea-air CO_2 flux (green) and from the pCO_2 products estimates (blue) for (a) the global ocean, and regionally for (b) the Atlantic Ocean, (c) the Pacific Ocean, (d) the Indian Ocean, (e) the Arctic Ocean, and (f) the Southern Ocean are shown. The shading indicates the uncertainty estimated as the respective standard deviation across all GOBMs and pCO_2 products. The uncertainty of the GOBM-estimate does not include the uncertainty of the riverine adjustment.

increasing trend is similarly represented by the two classes of estimates. Still, the difference of the long-term means of 0.20 ± 0.32 Pg C year^{-1} (12 ± 20% of the mean pCO_2-product estimate) is non-negligible. Moreover, the difference of annual mean fluxes between GOBMs and pCO_2-products varies with time, with the GOBMs exceeding 20% of the average value of the pCO_2-products from 1985 to 1990, in 2009 and 2010, and from 2016 to 2018 (Figure 1a). Furthermore, the individual GOBM estimates within the model ensemble also differ substantially with an inter-model range of all GOBMs of 0.24 Pg C year^{-1} representing ∼17% of their average CO_2-flux. Even larger differences are found on the regional scale (Figures 1b–1f).

Regionally, the time-averaged F^{net} from 1985 to 2018 based on GOBMs and pCO_2-products agree well in the Pacific Ocean, the Indian, the Arctic Ocean, and the Southern Ocean (Table 1, Figure 1). However, in the Atlantic Ocean the GOBMs indicate a substantially smaller uptake than the pCO_2 products (Table 1, Figure 1b). The difference in the Atlantic Ocean starts to increase around the year 2000, the same time when the F^{net} estimates in the Arctic Ocean also start to diverge (Figure 1e). Furthermore, the GOBMs and the pCO_2 products do not show

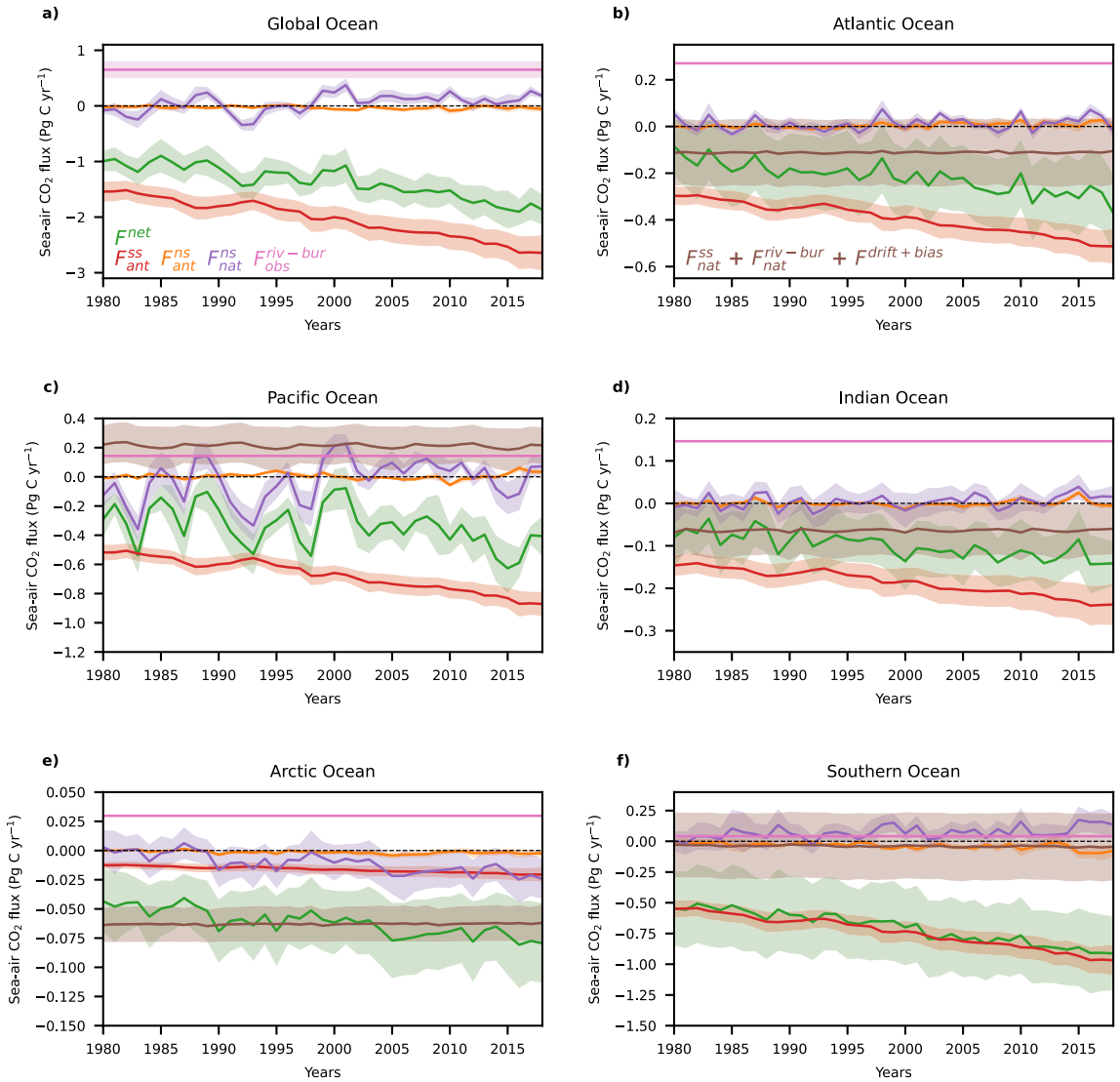

Figure 2. *Time series of sea-air CO_2 flux components globally and regionally from 1980 to 2018 based on global ocean biogeochemistry models (GOBMs).* The net sea-air CO_2 flux (F^{net}) integrated over each basin (green) and adjusted for $F_{obs}^{riv\text{-}bur}$ (pink), and the individual flux components from the GOBMs (F_{ant}^{ss} in red, F_{ant}^{ns} in orange, F_{nat}^{ns} in purple, and the sum of F_{nat}^{ss}, $F_{nat}^{riv\text{-}bur}$, and $F^{drift+bias}$ in brown) are shown for (a) the global ocean and regionally for (b) the Atlantic Ocean, (c) the Pacific Ocean, (d) the Indian Ocean, (e) the Arctic Ocean, and (f) the Southern Ocean. The shading indicates the respective standard deviation across all GOBMs. The uncertainty of F^{net} does not include the uncertainty of the riverine adjustment. The uncertainty of $F_{obs}^{riv\text{-}bur}$ is only quantified for the global ocean such that no uncertainties of the regional $F_{obs}^{riv\text{-}bur}$ can be shown.

the same decadal variability of F^{net} in the Southern Ocean (Figure 1f). The inter-model ensemble spread of simulated F^{net} is in absolute numbers largest in the Southern Ocean (~42% of the average CO_2-flux for 1985 to 2018), directly followed by the Atlantic Ocean (~67% of the average CO_2-flux for 1985 to 2018). A separation of F^{net} into its different flux components (see Section 2.2.3) allows us to identify the fluxes that are causing the inter-model differences. Globally, the largest contribution to the model spread in F^{net} in GOBMs stems from F_{ant}^{ss} (Figure 2a, Table S1 in Supporting Information S1). Regionally, the model spread in F^{net} is dominated by the model spread in the sum of F_{nat}^{ss}, $F_{nat}^{riv\text{-}bur}$, and $F^{drift+bias}$ in all basins but the Arctic Ocean (Figures 2b–2f, Table S1 in Supporting Information S1). The second largest contributions to the model spread are F_{ant}^{ss} and F_{nat}^{ns}. In the Arctic Ocean, the model spread of the sum of F_{nat}^{ss}, $F_{nat}^{riv\text{-}bur}$, and $F^{drift+bias}$ and the model spread in F_{nat}^{ns} are of similar size (Figure 2e, Table S1 in Supporting Information S1). The relatively large importance of F_{nat}^{ns} in the Arctic Ocean is mostly caused by sea ice decline, which is well represented in GOBMs and hence does not explain the model spread in F_{nat}^{ns}. Instead, the model spread in F_{nat}^{ns} is caused by the inter-model differences in simulated pCO_2 under the melting sea ice (Yasunaka et al., 2023).

Table 2
Global Ocean Carbon Fluxes (Pg C Year^{-1}) Averaged From 1980 to 2018

	Land-sea river carbon flux	Burial in sediments	$F_{nat}^{riv\text{-}bur}$	F^{SimB}	$F^{drift+bias}$
CESM-ETHZ	0.33	0.25	0.08	0.00	−0.08
CNRM-ESM2-1	0.61	0.74	−0.13	−0.14	−0.01
EC-Earth3	0.61	0.47	0.14	0.25	0.11
FESOM-REcoM-LR	0.00	0.00	0.00	−0.35	−0.35
MOM6-Princeton	0.18	0.10	0.08	−0.23	−0.31
MRI-ESM2-0	0.00	0.00	0.00	0.17	0.17
NorESM-OC1.2	0.00	0.54	0.00	0.00	0.00
ORCA025-GEOMAR	0.00	0.34	−0.34	−0.36	−0.02
ORCA1-LIM3-PISCES	0.61	0.59	0.02	−0.26	−0.28
GOBM-ensemble	0.27 ± 0.30	0.37 ± 0.27	−0.03 ± 0.15	−0.09 ± 0.23	−0.06 ± 0.18

Note. Positive fluxes indicate fluxes out of the ocean, except for the land-sea river carbon fluxes. $F_{nat}^{riv\text{-}bur}$ was estimated as the difference between the land-sea river carbon flux and the burial in sediments, except for NorESM-OC1.2. $F^{drift+bias}$ was derived as the difference between F^{SimB} and $F_{nat}^{riv\text{-}bur}$. The GOBM-ensemble values exclude MOM6-Princeton (see Section 2.1).

In the following sections, we will present and discuss the different flux components one by one across the GOBMs ensemble, assess how well they can be quantified by each of the hindcast simulations, identify reasons for mismatches between individual models and between GOBMs and pCO_2 products estimates, and propose recommendations for the set-up of GOBMs in future studies. A special focus will lie on the Atlantic Ocean, where the long-term mean difference between GOBMs and pCO_2 products estimates is largest, and on the Southern Ocean, where the various GOBM estimates differ the most and where the decadal variability of the difference between GOBMs and pCO_2 products is largest.

3.1. Sea-Air CO_2 Fluxes in the Steady State Control Simulation

3.1.1. Carbon Fluxes From Rivers and Into Sediments

The input of riverine carbon F_{nat}^{riv} and the sedimentation of carbon F_{nat}^{bur} is treated in various ways across the ensemble of GOBMs and varies from 0.00 to 0.61 Pg C year^{-1} and from 0.00 to 0.74 Pg C year^{-1}, respectively (Table 2). The difference between F_{nat}^{riv} and F_{nat}^{bur} varies between +0.14 and −0.54 Pg C year^{-1} and is −0.10 ± 0.23 Pg C year^{-1} when averaged over the eight GOBMs that provide all four simulations.

Here, we estimate $F_{nat}^{riv\text{-}bur}$ in each model to get a first-order estimate of the magnitude of these fluxes in the here considered GOBMs. For this estimation, we use the first-order assumption that $F_{nat}^{riv\text{-}bur} = F_{nat}^{riv} - F_{nat}^{bur}$ for all GOBMs except NorESM-OC1.2 (Table 2). This assumption ignores the potential influence of riverine and sediment fluxes of alkalinity, which change ocean pCO_2 and in consequence the uptake of carbon from the atmosphere, and nutrients, which change pCO_2 via primary production and remineralization (Gao et al., 2023; Terhaar, Orr, Ethé, et al., 2019). Additional information from NorESM-OC1.2 can be used here to demonstrate, as an example, the effect of alkalinity fluxes on $F_{nat}^{riv\text{-}bur}$. In NorESM-OC1.2, a larger carbon burial flux than carbon riverine flux does not lead to an uptake of carbon from the atmosphere because the burial of carbon is accompanied by a burial of alkalinity of similar size. The burial of alkalinity reduces the DIC storage capacity of the ocean. Overall, the alkalinity and carbon burial fluxes in NorESM-OC1.2 influence the sea-air CO_2 flux in similar magnitude but with opposite signs so that $F_{nat}^{riv\text{-}bur}$ is almost zero (Table 2). Similar information was not available for the other GOBMs. With the adjusted $F_{nat}^{riv\text{-}bur}$ for NorESM-OC1.2, the multi-model mean $F_{nat}^{riv\text{-}bur}$ is −0.03 ± 0.15 Pg C year^{-1}. In comparison, the model spread associated with F^{net} is 0.24 Pg C year^{-1}.

According to our approximation, none of the GOBMs simulates a resulting riverine and sediment flux-driven $F_{nat}^{riv\text{-}bur}$ comparable to the observation-based $F_{obs}^{riv\text{-}bur}$ of −0.65 ± 0.15 Pg C year^{-1} (Regnier et al., 2022). This mismatch is due to low land-sea riverine carbon fluxes in the GOBMs (0.27 ± 0.30 Pg C year^{-1}) compared to the observation-based estimates (0.85 ± 0.10 Pg C year^{-1} (Regnier et al., 2022)) and an on average high carbon burial in sediments by the GOBMs (0.37 ± 0.27 Pg C year^{-1}, yet this includes two GOBMs with no carbon burial

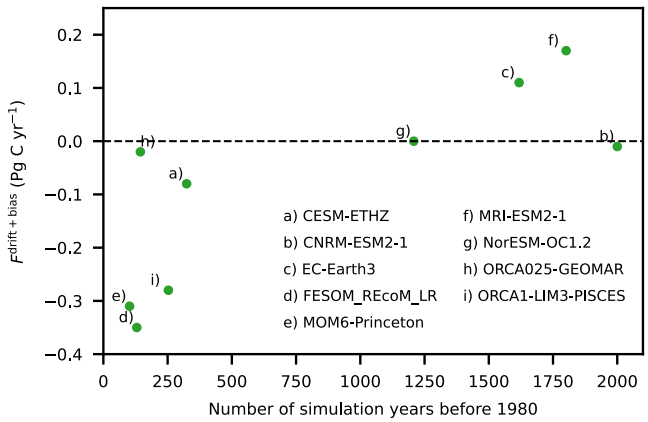

Figure 3. *Estimated bias and drift of global sea-air CO_2 fluxes related to the models not being in steady-state for nine global ocean biogeochemistry models against the length of their spin-up*. The length of the spin-up is defined as the number of simulated years at that resolution before the start of the analyzed period in 1980, while the bias and drift of the CO_2 flux ($F^{drift+bias}$) is determined as specified in Section 3.1.2, Table 2. As ORCA025-GEOMAR was branched from a previous spin-up from the same model but with a coarser resolution, the number of years shown here for ORCA025 is the sum of both spin-ups.

as seen in Table 2) compared to the observation-based estimate of 0.15 ± 0.03 Pg C $year^{-1}$ (Regnier et al., 2022). Global GOBM-estimates of F^{net} in this study are however unaffected by $F_{nat}^{riv\text{-}bur}$ (see Equation 9) as the simulated $F_{nat}^{riv\text{-}bur}$ is removed when subtracting F^{SimB} from F^{SimA} and replaced by the observation-based estimate of riverine and sediment fluxes. Yet, the regional $F_{nat}^{riv\text{-}bur}$ has to be accepted as an inherent uncertainty of each model as the regional $F_{nat}^{riv\text{-}bur}$ cannot be quantified across the GOBMs within RECCAP2. $F_{nat}^{riv\text{-}bur}$ may substantially affect the regional estimates of F^{net} and F_{nat}^{ss}, especially given the large relative size of $F_{obs}^{riv\text{-}bur}$ compared to the other flux components (see Figure 2).

3.1.2. Bias and Drift in the Sea-Air CO_2 Flux Due To Incomplete Spin-Up

Across the ensemble of GOBMs, the approximated global $F^{drift+bias}$, quantified as the difference between F^{SimB} and our estimation of $F_{nat}^{riv\text{-}bur}$ (Equation 3, Table 2), varies from −0.35 to 0.17 Pg C $year^{-1}$, with an ensemble mean of −0.06 ± 0.18 Pg C $year^{-1}$. The model spread around $F^{drift+bias}$ is of similar order as the model spread associated with the global F^{net} (0.24 Pg C $year^{-1}$). We assume that this is mostly a consequence of a too short spin-up and hence of models not being in a steady state. Here, the drift component in the sea-air CO_2 flux from 1980 to 2018 (calculated as the trend of the global air-sea CO_2 flux in simulation B) is less than ±0.002 Pg C $year^{-1}$ for all GOBMs (Hauck et al., 2020). Although our estimation of $F^{drift+bias}$ is uncertain due to several approximations in our methodology, it gives a first indication of the importance of the non-steady state for the model spread. A comparison between the number of simulated years before the start of the analysis period of each GOBM and the $F^{drift+bias}$ (Figure 3) suggests that a short spin-up is often insufficient to reduce $F^{drift+bias}$. A sufficiently long spin-up in each model to reach steady state may thus narrow down inter-model differences of regional F_{nat}^{ss} and F^{net}.

While $F^{drift+bias}$ does not directly affect our estimate of the global F^{net} (based on Equation 9), a GOBM not being in steady-state owing to an insufficient spin-up also affects temperature, salinity, DIC, alkalinity, and the circulation or chemical uptake capacity and may still affect global F^{net}. Yet it is unclear if these variables would have a larger or smaller bias due to an insufficient spin-up as a model's steady state may have large biases (Seferian et al., 2016). Regionally, $F^{drift+bias}$ directly affects F^{net} (see Section 2.2.3 and Equation 10).

3.1.3. Steady State Natural Sea-Air CO_2 Flux

The mean F^{SimB} estimates of the GOBMs from 1980 to 2018 (Figure 2) are −0.11 ± 0.14 Pg C $year^{-1}$ for the Atlantic Ocean, 0.21 ± 0.13 Pg C $year^{-1}$ for the Pacific Ocean, −0.06 ± 0.06 Pg C $year^{-1}$ for the Indian Ocean, and −0.06 ± 0.01 Pg C $year^{-1}$ for the Arctic Ocean. In the Southern Ocean, the F^{SimB} estimate of −0.04 ± 0.27 Pg C $year^{-1}$ of the GOBMs is twice as uncertain as in the other basins (Table S1 in Supporting Information S1). The relatively large uncertainty in the Southern Ocean may partly be the result of large inter-model differences in the simulated F_{nat}^{ss} fluxes, as dynamically complex regions like the Southern Ocean are difficult to simulate (Hauck, Gregor, et al., 2023; Sallée et al., 2013). Inter-model differences in $F^{drift+bias}$ likely also play a role for the uncertain F^{SimB} estimate as the Southern Ocean is the region where most of the oldest water masses are upwelled to the ocean surface (Caldeira & Duffy, 2000), which have not been in contact with the atmosphere during the spin up and would hence presumably cause a larger disequilibrium and a larger $F^{drift+bias}$ than in other ocean basins with less upwelling. The Southern hemisphere and especially the Southern Ocean are also the locations where the $F^{drift+bias}$ tends to be largest in ESMs (Séférian et al., 2016).

The GOBM-based estimates of F_{nat}^{ss} can be compared to inverse estimates of F_{nat}^{ss} (Mikaloff Fletcher et al., 2007) (see also Section 2.3). These inverse estimates of F_{nat}^{ss} show larger uptake in the Atlantic (−0.24 ± 0.08 Pg C $year^{-1}$) and Pacific Ocean (−0.07 ± 0.14 Pg C $year^{-1}$), more outgassing in the Southern Ocean (0.44 ± 0.11 Pg C $year^{-1}$), and similar uptake in the Arctic (−0.02 ± 0.01 Pg C $year^{-1}$) and Indian Ocean (−0.12 ± 0.04 Pg C $year^{-1}$). The differences between our estimates and that of Mikaloff Fletcher et al. (2007) are

partly due to different basin-definitions. Most prominently, the inverse estimate considers all areas south of 44°S as the Southern Ocean, which is different from our definition of the Southern Ocean (Figure S1 in Supporting Information S1). When changing the northern boundary of the Southern Ocean to 44°S, the regional F_{nat}^{ss} of the GOBMs changes to 0.27 ± 0.19 Pg C year^{-1}, which is in better accordance but still 0.18 Pg C year^{-1} smaller than the mean inverse-based estimate of 0.44 ± 0.11 Pg C year^{-1} and both estimates do not overlap within their uncertainties.

Partly, the large differences between the GOBM-based estimate and the inverse estimates of F_{nat}^{ss} might be due to $F^{drift+bias}$. Although $F^{drift+bias}$ cannot be quantified directly, previous studies suggest that drift and bias related fluxes in ESMs are mainly located in the Southern Ocean (e.g., Séférian et al., 2016). In the here presented ensemble of GOBMs, we also find a significant relationship ($r^2 = 0.58$, $p = 0.02$) with a slope of 1.03 between F^{SimB} in the Southern Ocean and F^{SimB} globally (Figure S2a in Supporting Information S1). A similar relationship occurs in no other ocean basin (Figures S2b–S2e in Supporting Information S1). However, the relationships between global and regional F^{SimB} are not only influenced by $F^{drift+bias}$ but also influenced by global and regional $F_{nat}^{riv-bur}$, which may disguise a relationship between regional and global F^{SimB} in basins where differences in riverine influx of carbon are large across the GOBM ensemble. As we are not able to regionally estimate $F_{nat}^{riv-bur}$, we hence can not be certain that $F^{drift+bias}$ is indeed stemming from the Southern Ocean. Nevertheless, if we assume that it is and adjusted the GOBM-based estimate of F_{nat}^{ss} in the Southern Ocean for this $F^{drift+bias}$, then the GOBM-based estimate of F_{nat}^{ss} would get closer to the inverse estimate of F_{nat}^{ss}.

3.2. Non-Steady State Natural Sea-Air CO_2 Flux

Averaged between 1980 and 2018, the GOBMs simulate a global F_{nat}^{ns} of 0.05 ± 0.05 Pg C year^{-1}. We further separated the inter-annual and decadal variability from the long-term signal by removing its linear trend (see e.g., DeVries, 2022). The simulated long-term signal shows a global F_{nat}^{ns} increase from 1980 to 2018 at a rate of 0.07 ± 0.02 Pg C year^{-1} decade^{-1} (Figures 2 and 4a). The tropical Pacific and the Indian section of the Southern Ocean are the main contributors to the trend toward stronger F_{nat}^{ns} carbon outgassing (Figures 2 and 4a). The average trend toward stronger outgassing of F_{nat}^{ns} is to a small part compensated by a trend toward non-steady uptake of natural CO_2 in the Northern Pacific and the Arctic Ocean (Figures 2e and 4a; Yasunaka et al., 2023). Across the model ensemble, large inter-model differences in the mean F_{nat}^{ns} flux exist in the tropical Southern Ocean, the sea ice edge in the North Atlantic and Arctic Ocean, and the eastern coastal upwelling systems (Figure 4b).

The globally simulated inter-annual and decadal variability in F_{nat}^{ns} of 0.16 ± 0.03 Pg C year^{-1} is similar across the GOBMs (Figure 2a), likely because many models use the same atmospheric reanalysis products for their forcing. Most of the inter-annual variability in F_{nat}^{ns} occurs in the tropical Pacific Ocean and the high-latitude oceans (Figure 4c). Though the pattern of variability is similar across the GOBMs, relatively large inter-model differences are found in the Southern Ocean, north-western Pacific Ocean, the North Atlantic subpolar gyre, and the Peruvian upwelling system (Figure 4d). The inter-annual and decadal variability in F_{nat}^{ns} is the dominant contributor to the inter-annual and decadal variability of F^{net} in GOBMs and is globally six times larger than the variability in the climate-driven variability in the anthropogenic sea-air CO_2 fluxes (F_{ant}^{ns}) and regionally two to six times larger (Figure 2). The simulated temporal variability of F^{net} in the Pacific Ocean is driven by F_{nat}^{ns} (Figure 2c) and resembles the variability of F^{net} in the pCO_2 products (Figure 1). This good agreement indicates that the GOBMs represent the dominant source of Pacific sea-air CO_2 flux variability, El-Niño and La-Niña (Feely et al., 1999), well.

3.3. Anthropogenic Carbon Fluxes and Storage

3.3.1. Interior Ocean Anthropogenic Carbon Storage

The spatial distribution of the interior ocean C_{ant}-storage since the beginning of the industrial period simulated by the here analyzed GOBM ensemble resembles that of the TTD- and ΔC^*-estimate (Figure 5, Figure S3 in Supporting Information S1) and that of other observation- and model-based studies (e.g., Davila et al., 2022; Khatiwala et al., 2013). The salinity biases of PlankTOM12 led to an anthropogenic carbon storage pattern that does not resemble any of the observation-based estimates and led to its exclusion from all GOBM results in terms of multi-model mean and standard deviation (Text S1 in Supporting Information S1). While the TTD- and ΔC^*-based estimates and the here analyzed eight GOBMs agree that the largest accumulation of C_{ant} per surface area is

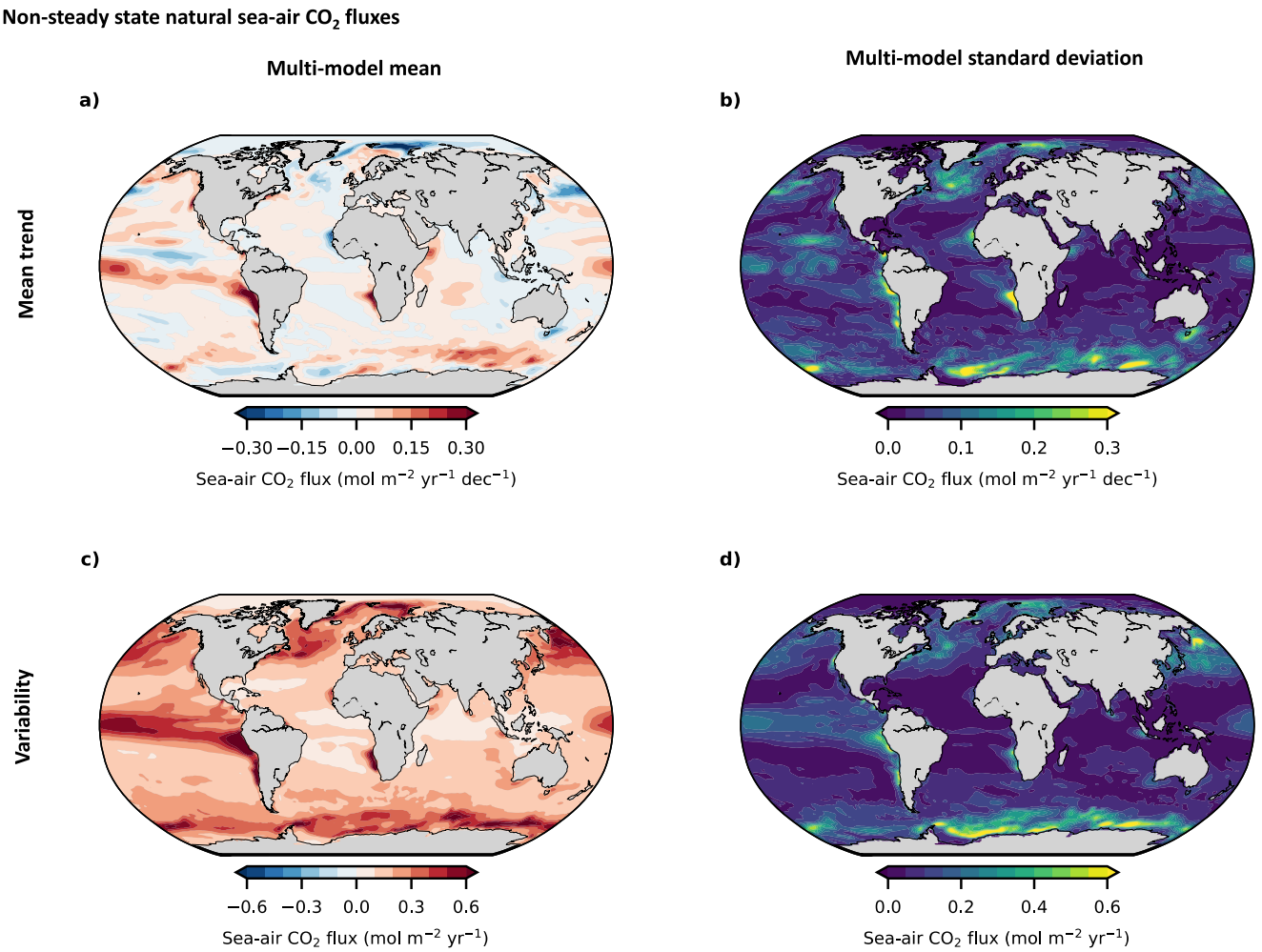

Figure 4. *Non-steady state natural sea-air CO_2 fluxes for eight global ocean biogeochemistry models*. Maps of the multi-model (a) mean trend and (b) standard deviation of the trend of the natural non-steady state sea-air CO_2 flux from 1980 to 2018, as well as maps of the multi model (c) mean inter-annual variability and (d) standard deviation of the inter-annual variability of the natural non-steady state sea-air CO_2 flux (linear trend is removed).

located in the North Atlantic and at the northern limit of the Southern Ocean around 45°S, the inter-model spread is high in these regions.

When only integrating over cells where estimates from associated observation-based products exist (see Section 2.3), the GOBM ensemble underestimates the integrated interior C_{ant} from surface to 3,000 m depth that accumulated since preindustrial times. The simulated multi-model mean interior ocean C_{ant} is 83 ± 15 Pg C in 1994, 22% (23 Pg C) lower than the ΔC^*-estimate, and 102 ± 12 Pg C in 2002, 30% (44 Pg C) lower than the TTD-estimate. Most prominent differences are in the North Atlantic and Southern Ocean (Figure 5). These differences may be caused by the starting dates of the GOBM simulations that vary between 1765 and 1870 (see Section 3.3.2) and biases in GOBM dynamics and biogeochemistry (see Sections 3.3.3.1 and 3.3.3.2). In addition, the TTD-estimate might be biased high in the Southern Ocean and the North Atlantic due to its methodology (DeVries, 2014; Matear et al., 2003; Terhaar, Tanhua, et al., 2020; Waugh et al., 2006) and the ΔC^*-methodology might lead to an overestimation of C_{ant} in the upper water column and a negative bias in deeper waters (Matsumoto & Gruber, 2005).

As for the C_{ant}-storage since 1800, the spatial pattern of the simulated interior ocean C_{ant}-storage changes from 1994 to 2007 of the GOBMs resembles that of the eMLR(C^*)-estimate (Figure 5, Figure S4 in Supporting Information S1). Over this recent period, the global model mean C_{ant}-storage change of 25 ± 3 Pg C (only integrating over cells where C_{ant} estimates from the eMLR(C^*) method exist) is also smaller than the eMLR(C^*)-

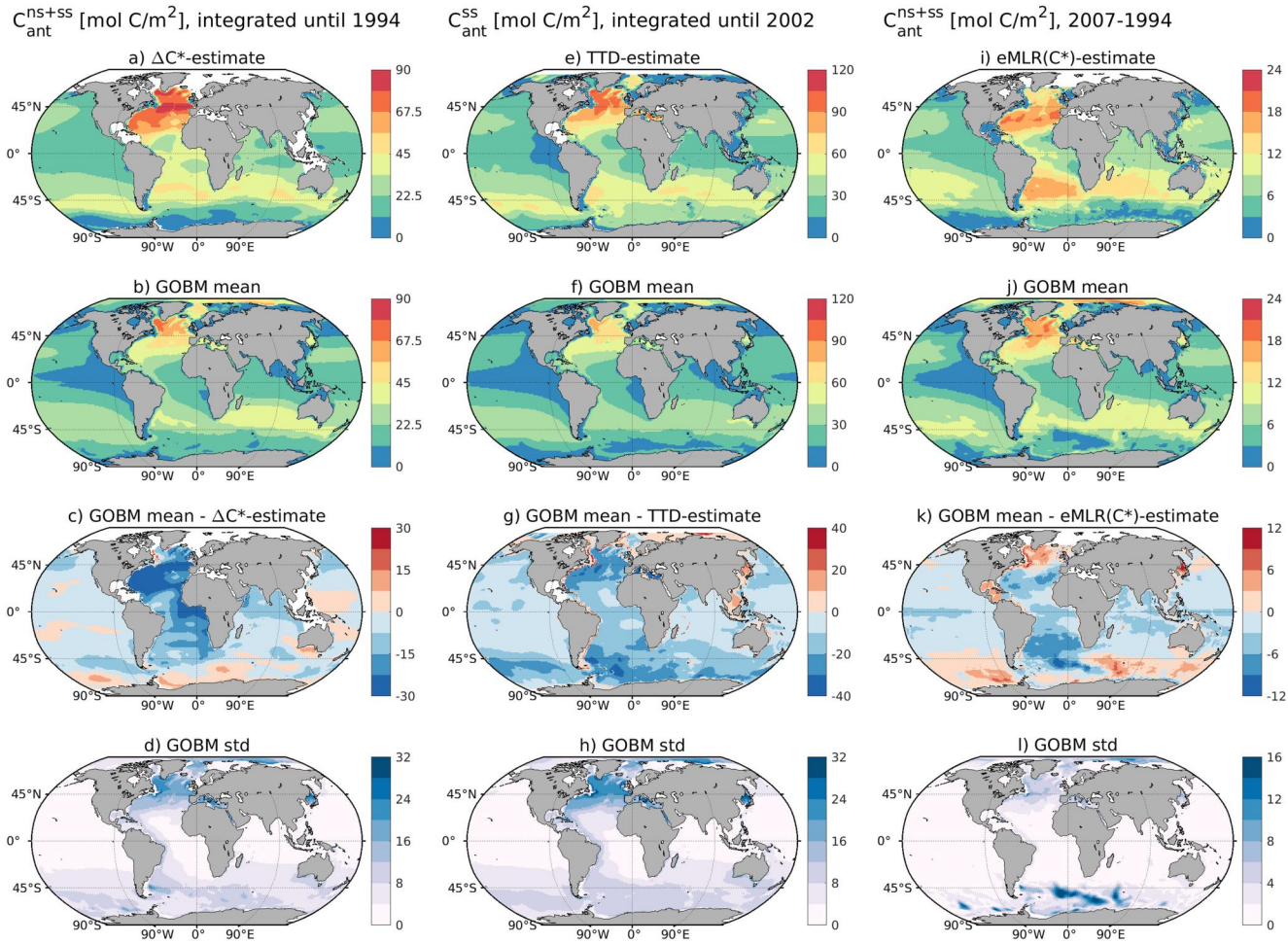

Figure 5. *Column inventories of historic and contemporary anthropogenic carbon storage changes, integrated from surface to 3,000 m depth*. Visualized are (a, e, i) observation-based estimates and related model-estimates based on eight global ocean biogeochemistry models (GOBMs), shown as (b, f, j) model mean, (c, g, k) difference between model-mean and observation-based estimates and (d, h, l) multi-model standard deviation. Panels (a–d) show results for C_{ant}^{ns+ss} from the ΔC^*-estimate for the period 1800–1994 and GOBM estimates from start date of each simulation to 1994, (e–h) show results for C_{ant}^{ss} from the TTD-estimate for the period 1800–2002 and GOBM estimates from start date of each simulation to 2002, while panels (i–l) show results for C_{ant}^{ns+ss} from 1994 to 2007, contrasting the eMLR(C^*)-estimate with the GOBM estimates. Individual results for each of the considered GOBMs and PlankTOM12 are presented in Figures S3 and S4 in Supporting Information S1.

estimate, but only by approximately 20% (6 Pg C). The underestimation of the contemporary C_{ant}-storage change by GOBMs is likely smaller than the underestimation of C_{ant}-storage changes since 1800 because the later starting date of several GOBMs (Section 3.3.2) has a smaller effect on contemporary C_{ant}-storage changes. Regionally, differences between the GOBM mean and the eMLR(C^*)-estimate (Figure 5) are most prominent in the Atlantic and Southern Ocean (Hauck, Gregor, et al., 2023; Hauck, Nissen, et al., 2023). The eMLR(C^*)-estimate indicates an anomalously high rate of C_{ant}-change in the South Atlantic for the period from 1994 to 2007 and an anomalously low rate of C_{ant}-change in the subpolar North Atlantic and the Indian and Pacific sectors of the Southern Ocean (Gruber, Clement, et al., 2019), which was attributed to a temporary slow-down and reorganization of the North Atlantic overturning circulation (Fröb et al., 2016; Pérez et al., 2013; Steinfeldt et al., 2009) and changes in the Southern Ocean meridional overturning circulation and ventilation of water masses (Tanhua et al., 2017; Waugh et al., 2013). The GOBMs do not exhibit the regionally anomalous accumulation of C_{ant} that is apparent in the eMLR(C^*)-estimate so that the GOBM ensemble mean is smaller than the eMLR(C^*)-estimate in the South Atlantic and subtropical North Atlantic and larger than the eMLR(C^*)-estimate in the subpolar North Atlantic and the Indian and Pacific sectors of the Southern Ocean (Hauck, Gregor, et al., 2023; Hauck, Nissen, et al., 2023).

However, the eMLR(C^*)-estimate might also overestimate the strength of these anomalies, due to structural biases in the reconstructed changes of C_{ant} (Clement & Gruber, 2018; Gruber, Clement, et al., 2019).

Overall, the comparison of simulated and observation-based C_{ant} confirms that the GOBMs underestimate the oceanic storage of anthropogenic carbon and hence F_{ant}^{ss} by 20%–30% as suggested by the Global Carbon Budget (Friedlingstein et al., 2022). Moreover, across the GOBM ensemble there exists a strong relationship between the simulated C_{ant} storage in 1994 since the beginning of the industrialization and the simulated change in C_{ant} storage from 1994 to 2007 across the model ensemble (Figure S5 in Supporting Information S1) suggesting a bias in the model mean state that persists over centuries. In the following sections, we will analyze the model setups, and simulated circulation and biogeochemistry to identify reasons for the underestimation of F_{ant}^{ss} by the GOBM ensemble.

3.3.2. Influence of Pre-Industrial Atmospheric CO_2 Mixing Ratio and Simulation Starting Date on Anthropogenic Carbon Uptake

The difference in the simulated sea-air CO_2 flux from 1980 to 2018 between simulations starting in 1765 and those starting in 1850 is simulated by the EMIC Bern3D-LPX to be 0.04–0.06 Pg C year^{-1} (Figure 6d), depending on the ocean mixing strength (Figure 6c) (see Section 2.6 for details of this setup). From 1765 to 1995, the difference in the simulated cumulative sea-air CO_2 flux due to the late starting date is 18.2–22.7 Pg C, around 15%–19% of the cumulative anthropogenic carbon uptake from 1765 to 1995 (Sabine et al., 2004). Slightly less than half of this difference (8.4–9.0 Pg C) occurs before 1850 and slightly more than half of this difference (9.8–13.7 Pg C) occurs after 1850. The difference after 1850 is caused by the difference in atmospheric pCO_2 in the respective spin-up and pre-industrial control simulations, which is equal to the atmospheric starting year of the simulation. As the atmospheric pCO_2 in 1765 was lower than in 1850, all water masses in the simulation that started in 1765 also have a lower pCO_2 at the beginning of the simulation than the water masses in the simulations that started in 1850. When water masses upwell for the first time over the simulation period, they take up carbon and the water masses with a lower pCO_2 in the simulation starting in 1765 take up more carbon. For the REC-CAP2 time period (1980–2018), it is thus mostly the difference in atmospheric pCO_2 in the respective spin-up and pre-industrial control simulations that causes an underestimation of the carbon sink over that period. Therefore, most differences in the air-sea CO_2 flux from 1980 to 2018 occur in regions of strong upwelling, such as the Southern Ocean (Figure 6b).

In comparison, the two offline approaches by Bronselaer et al. (2017) estimate an underestimation of the ocean carbon sink of 28.7 ± 4.6 Pg C for the period from 1765 to 1995 when starting simulations in 1850 instead of 1765 and using the respective atmospheric pCO_2 concentrations from the starting years during the spin-up. Also in these offline estimates, more than 50% of this underestimation (~17 Pg C) is estimated to occur after 1850. Hence, Bronselaer et al. (2017) suggest a similar division of the adjustment before and after 1850, but their estimate for the entire period is around 40% larger than the estimate by Bern3D-LPX. A possible reason for the lower adjustment estimates by Bern3D-LPX may be the coarse resolution (40 × 41 horizontal cells and only three cells in the upper 126 m) leading likely to a more diffusive transport than in models with a higher horizontal resolution. A more diffusivity-driven tracer transport reduces the transport contribution from upwelling of older water masses to the surface and hence reduces the impact of these water masses, for example, in coastal upwelling systems where no difference in air-sea CO_2 flux is simulated in Bern3D-LPX between simulations starting in 1765 and 1850.

Thus, the adjustment simulated by Bern3D-LPX for the air-sea CO_2 flux from 1980 to 2018 of 0.04–0.06 Pg C year^{-1} might be underestimated by around 40%. Eventually, only GOBM simulations starting in 1765 allow quantifying the underestimation with certainty.

3.3.3. Steady-State Anthropogenic Sea-Air CO_2 Fluxes

The large-scale pattern of the steady-state anthropogenic sea-air CO_2 flux (F_{ant}^{ss}) averaged from 1980 to 2018 is similar across all GOBMs with the largest regional uptake rates in the high latitude North Atlantic and the Southern Ocean (Figure 7). The various numerical representations of the ocean circulation in the GOBMs result in a large model spread of F_{ant}^{ss} and C_{ant} in both North Atlantic and Southern Ocean (Figure 5, Section 3.3.1), similar to previous results from GOBMs (Orr et al., 2001) and ESMs (Frölicher et al., 2015; Goris et al., 2018; Terhaar, Frölicher, & Joos, 2021).

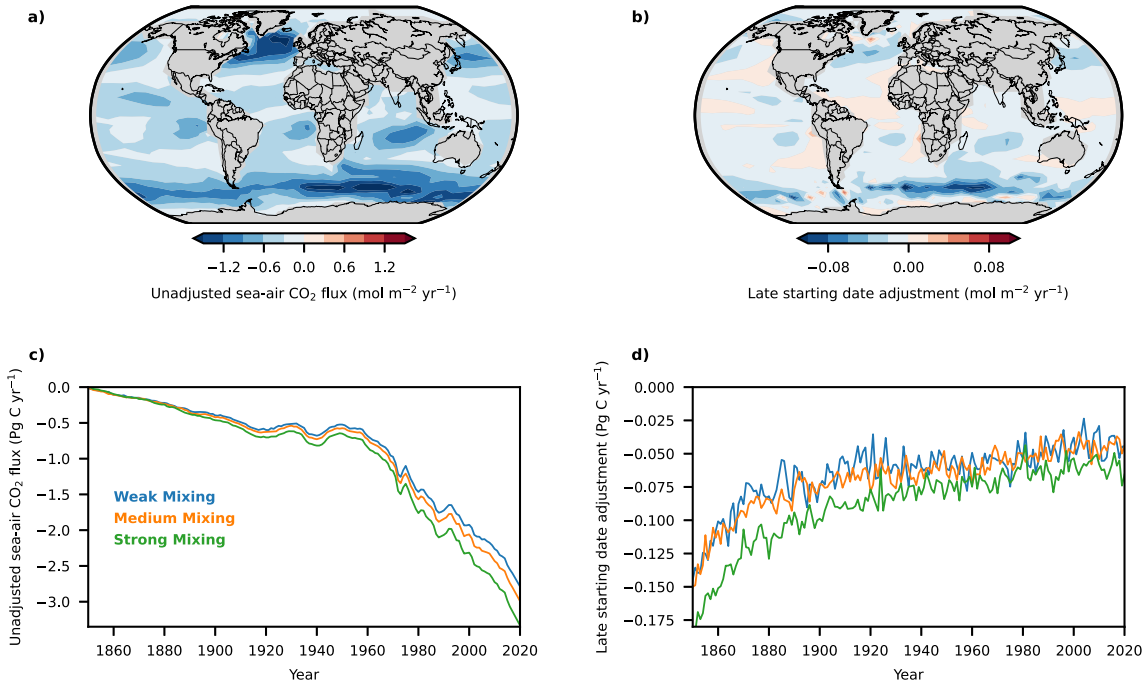

Figure 6. *Difference in anthropogenic sea-air CO_2 fluxes due to different starting dates in Bern3D-LPX*. Maps of (a) the anthropogenic sea-air CO_2 flux (steady-state and non-steady state) averaged from 1980 to 2018 and averaged over 3 Bern3D-LPX simulations with varying ocean mixing that start in 1850 and (b) the difference of the same flux between the simulations that start in 1765 and those that start in 1850. Time series of (c) the anthropogenic sea-air CO_2 flux from simulations starting in 1850 with weak (blue), medium (orange), and strong (green) ocean mixing, and time series of (d) the difference in the anthropogenic sea-air CO_2 flux between simulations starting in 1850 and those that start in 1765 for the same ocean mixing strengths.

3.3.3.1. Role of Ocean Circulation on Steady State Anthropogenic Sea-Air CO_2-Fluxes in the Atlantic and the Southern Ocean

In the Atlantic Ocean, the AMOC is the underlying driver of the uptake and storage of C_{ant}. It transports surface waters with high C_{ant} (Pérez et al., 2013) and subsurface waters with low C_{ant} (Ridge & McKinley, 2020) northwards. The subsurface waters outcrop in the subpolar gyre and are hence a sink of C_{ant} (Ridge & McKinley, 2020). Both water masses are eventually transformed into deep water and transported southward. The AMOC is also the main driver of F_{ant}^{ss} differences in the Atlantic across ensembles of ESMs from CMIP5 and

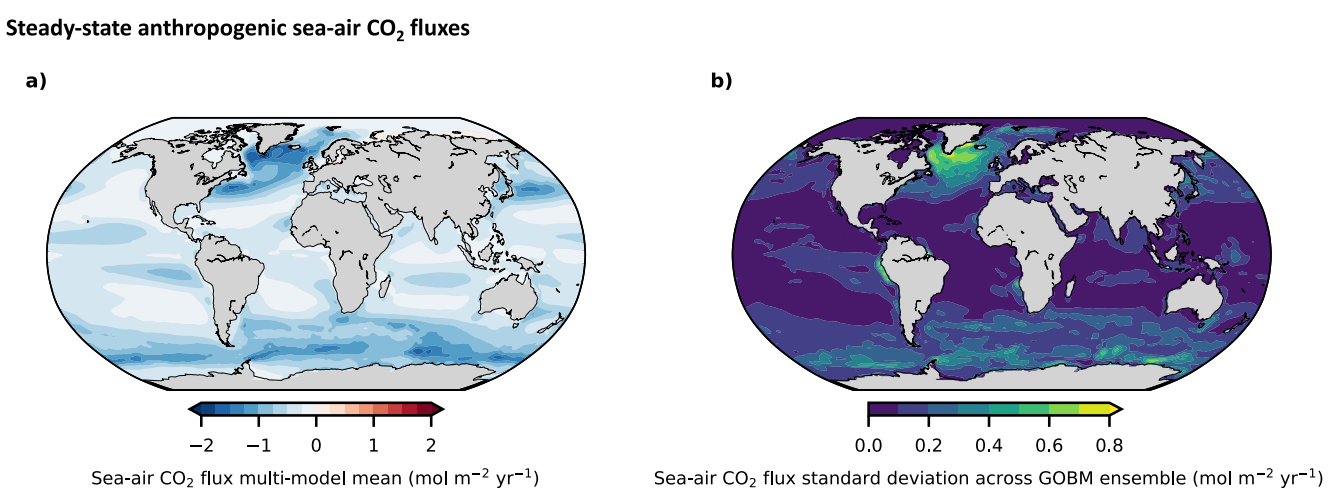

Figure 7. *Simulated model mean and model spread of the steady-state anthropogenic CO_2 flux*. Maps of (a) the multi-model mean and (b) multi-model standard deviation of the steady state anthropogenic sea-air CO_2 flux averaged from 1980 to 2018 for eight global ocean biogeochemistry models (GOBMs).

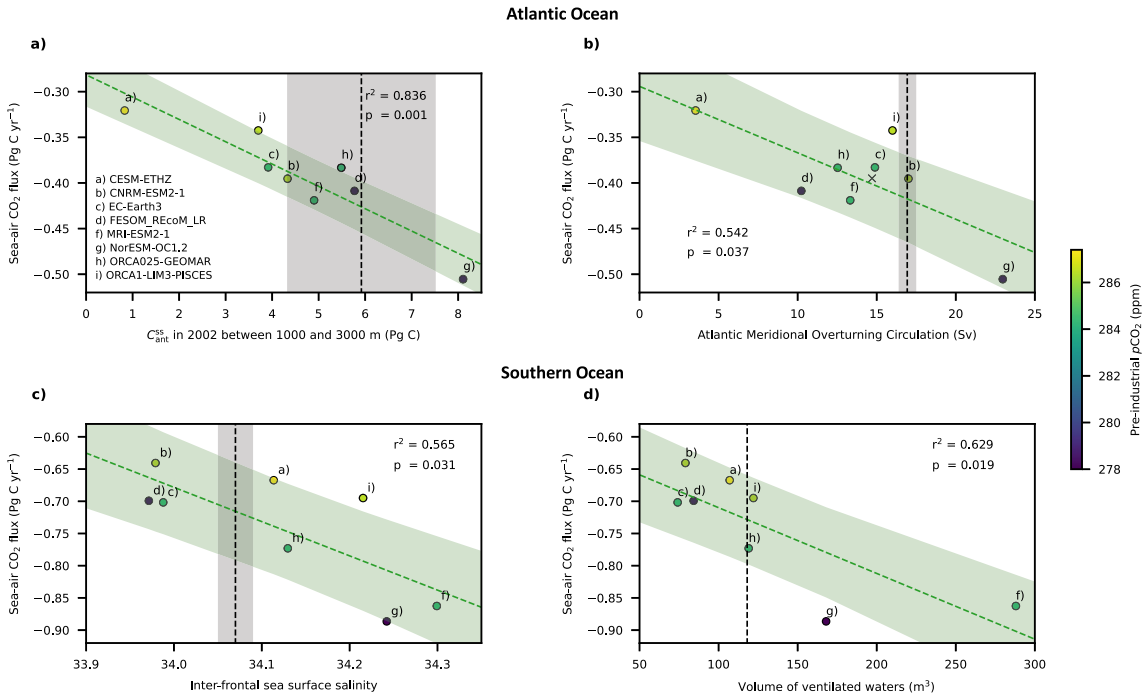

Figure 8. *Constrained steady-state anthropogenic carbon uptake in the Atlantic and Southern Ocean*. Steady-state anthropogenic carbon uptake averaged from 1980 to 2018 of (a, b) the Atlantic and (c, d) the Southern Ocean, plotted against (a) the Atlantic steady-state anthropogenic carbon storage between 1,000 and 3,000 m depth for the year 2002, (b) the Atlantic Meridional Overturning Circulation at 26°N averaged from 2005 to 2018, (c) the inter-frontal sea surface salinity and (d) the volume of ventilated waters in the Southern Ocean. Linear fits (green dashed line) with 68% projection intervals (green shaded area) are shown across global ocean biogeochemistry models (GOBMs) (colored dots with a green outline). The colors of the dots indicate the pre-industrial atmospheric pCO_2 for each GOBM. Observation-based estimates and their uncertainties are marked with dashed black lines and black shaded areas (see Section 2.3 for a description of utilized observation-based estimates and their uncertainties). The cross in (b) indicates an additional simulation with CESM-ETHZ (see Section 2.1).

CMIP6 (Goris et al., 2023; Terhaar et al., 2022), linking F_{ant}^{ss} and the amount of C_{ant} that was transported below 1,000 m across these model ensembles (Goris et al., 2018, 2023).

Correlations between F_{ant}^{ss} and (a) the AMOC at 26.5°N or (b) the storage of C_{ant} between 1,000 and 3,000 m in the high latitude North Atlantic also occur across this ensemble of GOBMs and can be used to identify emergent constraints (Figures 8a and 8b). In combination with the respective observation-based estimates, the average annual Atlantic F_{ant}^{ss} from 1980 to 2018 can be constrained from -0.39 ± 0.05 to -0.43 ± 0.06 Pg C year^{-1} when using the deep ocean C_{ant}^{ss} storage and to -0.42 ± 0.05 Pg C year^{-1} when using the AMOC. The constraints identify a common bias in the GOBMs toward too small AMOC strengths (mean underestimation of 18%) and C_{ant}^{ss} storage below 1,000 m (mean underestimation of 22%), and hence Atlantic F_{ant}^{ss} (mean underestimation of 8%–10%, depending on the used constraint). In comparison, the CMIP6 ESMs also simulate a wide range of AMOCs but their multi-model mean is close to the observed values (Terhaar et al., 2022). Among the RECCAP2 GOBMs, a notable exception is the original version of CESM-ETHZ, as its AMOC is very small (<5 Sv). This was improved upon significantly in the new version, where an adjustment of the restoring timescale for salinity in the Southern Ocean increased the simulated AMOC to near 15 Sv. This highlights the high sensitivity of the AMOC to uncertainties in the buoyancy forcing, and especially the freshwater fluxes.

While the emergent constraints are able to correct for the underestimation of the AMOC across our GOBM ensemble, they cannot reduce the uncertainties around the Atlantic F_{ant}^{ss} estimate due to the relatively large uncertainty of the observation-based estimate in case of the C_{ant}^{ss} storage as well as the relatively weak but significant correlation between the AMOC and the Atlantic F_{ant}^{ss} ($r^2 = 0.54$, $p = 0.04$). This weak correlation may partly be driven by the varying starting dates as GOBMs with a later or earlier starting date tend to have smaller or higher F_{ant}^{ss} than expected from the fit, respectively (Figure 8b). The correlation of the C_{ant}^{ss} storage and F_{ant}^{ss} is stronger ($r^2 = 0.84$, $p = 0.001$) because both variables are more directly related to each other and coherently affected by diverse starting dates. The relationships between Atlantic F_{ant}^{ss} and (a) AMOC and (b) C_{ant}^{ss} storage

between 1,000 and 3,000 m in the high latitude North Atlantic stem from the North Atlantic, where the associated correlations are higher ($r^2 = 0.69$ for AMOC and $r^2 = 0.88$ for C_{ant}^{ss} storage).

In the Southern Ocean, the magnitude of F_{ant}^{ss} also depends sensitively on the overturning circulation (Caldeira & Duffy, 2000; Mignone et al., 2006; Sarmiento et al., 1992), consisting here of the upwelling of circumpolar deep water close to the polar front, which is mainly transported northward, transferred to mode and intermediate waters, and eventually subducted at the subtropical front below the light subtropical surface waters into the ocean interior (Marshall & Speer, 2012; Talley, 2013). Across two ensembles of ESMs, it could be demonstrated that the volume of ventilated mode and intermediate waters in the Southern Ocean is highly correlated with the sea surface density between the polar front and the subtropical front, that is, a higher sea surface density in the region of mode and intermediate water formation allows for more and deeper penetration of these water masses into the ocean interior and hence more F_{ant}^{ss} uptake (Terhaar, Frölicher, & Joos, 2021). As the density in the region of interest is almost entirely driven by the salinity (Supplement of Terhaar, Frölicher, and Joos (2021)), the sea surface salinity can be used as a proxy for sea surface density.

Our ensemble of GOBMs contains a similar range of inter-frontal sea surface salinities (~0.4) as the ESM ensemble and confirms the Southern Ocean relationships between F_{ant}^{ss} and (a) the inter-frontal sea surface salinity, that is, the mean surface salinity in the subtropical-polar frontal zone ($r^2 = 0.57$, $p = 0.03$), and (b) the volume of ventilated waters ($r^2 = 0.63$, $p = 0.03$) (Figures 8c and 8d). The constraint with the sea surface salinity as predictor reduces the magnitude of F_{ant}^{ss} in the Southern Ocean slightly from -0.74 ± 0.09 Pg C year^{-1} to -0.72 ± 0.08 Pg C year^{-1} (less uptake, 11% smaller uncertainty, Figure 8c). The relatively weak but significant correlation (compared to a correlation of $r^2 = 0.74$ for ESMs when considering the oceanic CO_2-uptake until 2005 (Terhaar, Frölicher, & Joos, 2021)) between the sea surface salinity and F_{ant}^{ss} can partly be explained by different starting dates as GOBMs with a late or early starting date have a smaller or larger absolute F_{ant}^{ss} than expected from the linear fit between the mean surface salinity in the subtropical-polar frontal zone, respectively (Figure 8c). A common starting date for all GOBMs, would likely have tightened the relationship of the emergent constraints using the AMOC and the interfrontal salinity, and decreased the uncertainty of the constrained estimate. We do not use the volume of ventilated waters to constrain F_{ant}^{ss} because the scarcity of subsurface observations would have resulted in large uncertainties of the observational constraint. As all GOBMs are forced with historical reanalysis data, the location of the fronts does not vary as much across the GOBM ensemble as it does for the ESM ensembles (Terhaar, Frölicher, & Joos, 2021). As the biomes are partly defined based on the location of these fronts, the biome-averaged sea surface salinity in the two Southern Ocean biomes north of the sea ice edge can also be used as a constraint for GOBMs (Hauck, Gregor, et al., 2023; Hauck, Nissen, et al., 2023). Some of the GOBMs also restore the salinity at the ocean surface toward observed salinities. Despite this restoring, some GOBMs still overestimate the inter-frontal salinity substantially (Figures 8c and 8d).

While the here considered emergent constraints change the average annual F_{ant}^{ss} from 1980 to 2018 in Atlantic and Southern Ocean only slightly, the associated relationships also allow to identify larger biases in F_{ant}^{ss} in individual models caused by circulation biases. Furthermore, the influence of circulation biases on F_{ant}^{ss} increases in magnitude with increasing atmospheric F_{ant}^{ss}. Therefore, the difference between constrained and unconstrained F_{ant}^{ss} increases over time (Figure S6 in Supporting Information S1) and a GOBM ensemble with circulation biases will have smaller trends in F_{ant}^{ss} and deviate from the true F_{ant}^{ss} with time.

3.3.3.2. Surface Ocean Carbonate Chemistry

The pCO_2/alkalinity products suggest that the largest chemical surface ocean uptake capacity (defined here as $\partial DIC/\partial[pCO_2]$, see Section 2.3) is found in the subtropical gyres, while the smallest chemical uptake capacities are in the polar oceans and the eastern tropical Pacific (Figure 9a). The GOBMs reproduce this pattern on average (Figure 9b) but show larger chemical uptake capacities in the tropical and subtropical oceans, and smaller chemical uptake capacities in the subpolar gyres, most of the Southern Ocean, the Labrador Sea, and the Arctic Ocean (Figure 9c). The inter-model variability is small in most places apart from sea ice regions in the Arctic Ocean and in eastern upwelling systems west of South America and Africa (Figure 9d), suggesting common biases in the chemical uptake capacities across the GOBM ensemble.

Globally, the chemical uptake capacity of the eight GOBMs is similar to that of the pCO_2/alkalinity products and of GLODAPv2 (Figure 9e). This capacity is directly linked to the surface alkalinity (Figure 9h) as GOBMs with a high buffer capacity have also high surface ocean CO_3^{2-} concentrations (Figure 9f), a high difference in surface

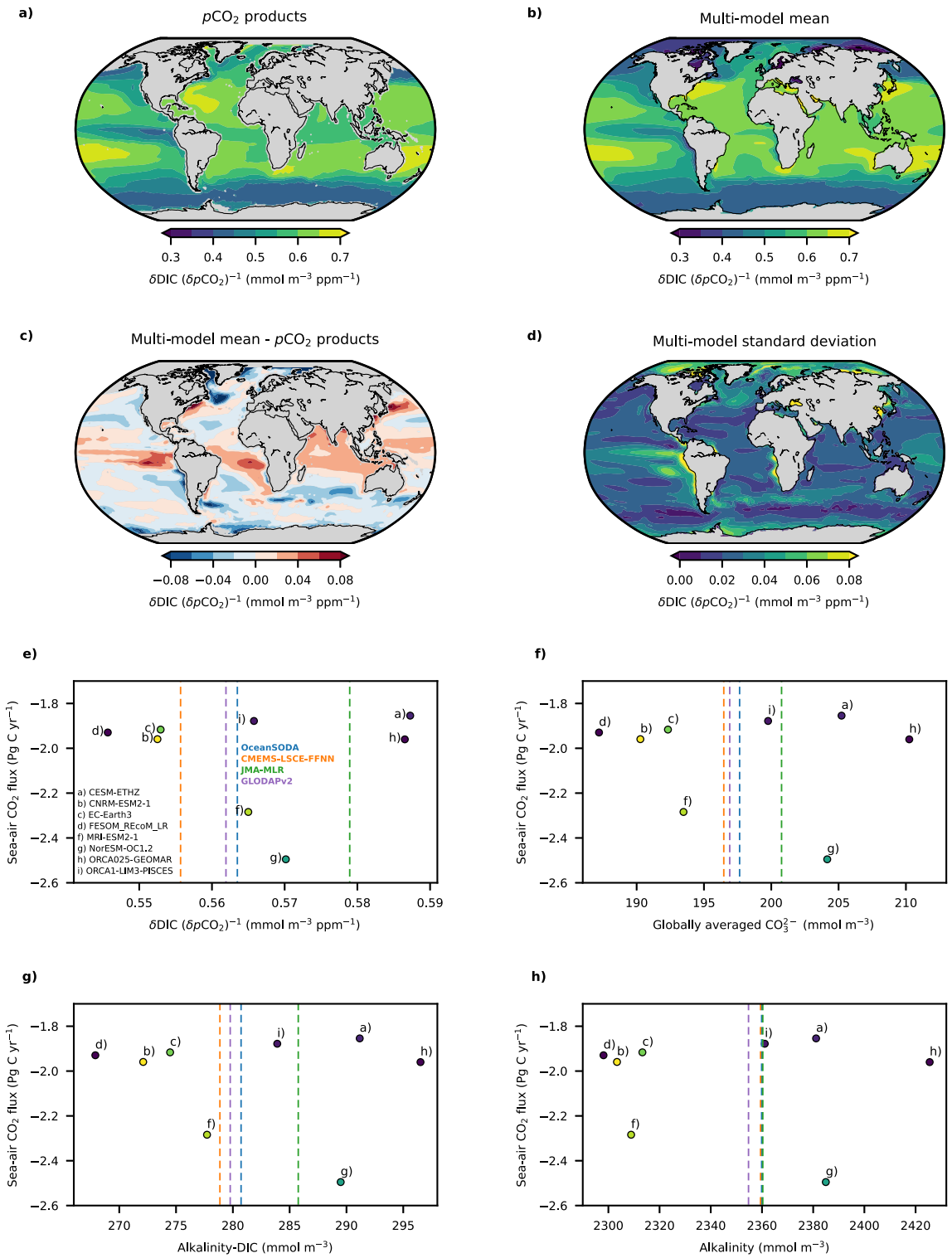

Figure 9. *Surface ocean chemical uptake capacity and its relationship to the steady-state anthropogenic sea-air CO_2 flux*. Maps of the increase in dissolved inorganic carbon (DIC) per increase in pCO_2 averaged from 1986 to 2018 based on (a) 3 pCO_2/alkalinity products (average of OceanSODA-ETHZ, CMEMS-LSCE-FFNN, and JMA-MLR) and (b) eight global ocean biogeochemistry models (GOBMs) (multi-model mean), as well as of (c) the difference between the pCO_2/alkalinity products mean and the GOBM multi-model mean and (d) the multi-model standard deviation. Scatterplots of temporal averages (1982–2018) of the accumulated global anthropogenic sea-air CO_2 flux against the global mean area-weighted (e) increase in DIC per increase in pCO_2, (f) surface ocean CO_3^{2-} concentration, (g) difference between surface ocean alkalinity and DIC, and (h) the global surface ocean alkalinity. The colors of each dot that represents a GOBM indicate the number of simulated years before the start of the analyzed period in 1980, and the dashed lines indicate each pCO_2/alkalinity product and GLODAPv2 (GLODAPv2-values are not used in panels (a) and (c) as they only represent the year 2002) for the variables on the respective *x*-axis.

ocean alkalinity and DIC (Sarmiento & Gruber, 2006) (Figure 9g) and high surface ocean alkalinity (Figure 9h). A similar relationship was also found across an ensemble of ESMs (Terhaar et al., 2022) and underlines the importance of alkalinity (Middelburg et al., 2020; Planchat et al., 2023).

We find that GOBMs represent surface ocean alkalinity better (range of ~2,300–2,425 mmol m^{-3}) than ESMs (range of 2,225–2,415 mmol m^{-3}, Terhaar et al. (2022)), potentially due to their atmospheric forcing from historical reanalysis and the use of salinity restoring toward observations, and hence a more realistic upwelling of circumpolar deep water with high alkalinity (Millero et al., 1998; Takahashi et al., 1981). Indeed, the GOBMs with the highest ventilation of surface waters in the Southern Ocean and hence also with the strongest upwelling of circumpolar deep waters with high alkalinity (MRI-ESM-2.0 and NorESM-OC1.2), are the GOBMs that show the highest chemical uptake capacity in the Southern Ocean (Figures S7 and S8 in Supporting Information S1).

For the GOBMs, their globally different chemical uptake capacities do not explain their global differences in F_{ant}^{ss} (Figure 9e), although studies with ESMs found such a relationship (Terhaar et al., 2022). Possible reasons for no emerging relationship between F_{ant}^{ss} and the chemical uptake capacity, CO_3^{2-}, or the alkalinity across the GOBM ensemble are differences in F_{ant}^{ss} due to ongoing $F^{drift+bias}$ or different starting dates of the simulations and corresponding differences in the atmospheric pCO_2 during the spin-up and pre-industrial control simulations (Section 3.3.2). If a GOBM has a large negative or positive $F^{drift+bias}$, its upwelling waters have too low or high DIC, too high or low CO_3^{2-}, and hence a chemical uptake capacity that is too high or low, respectively. With time, the additional surface ocean DIC from $F^{drift+bias}$ reduces the chemical uptake capacity so that it is effectively smaller than the one expected from the theoretical chemical uptake capacity. Thus, $F^{drift+bias}$ adds considerable noise so that a potential relationship between the chemical uptake capacity and F_{ant}^{ss} may not be identifiable. When considering only the four GOBMs with a longer spin-up than 1,000 years, a relationship indeed emerges (Figures 9e–9h).

3.4. Non-Steady State Anthropogenic Sea-Air CO_2 Flux

Globally, the GOBMs show an average F_{ant}^{ns} from 1980 to 2018 of -0.03 ± 0.04 Pg C $year^{-1}$ (Figure 10). As for F_{nat}^{ns}, we separate F_{ant}^{ns} into an interannual and decadal variability component and a long-term linear trend component. On average, the GOBMs simulate that the long-term trend increases the uptake of C_{ant} in the Southern Ocean and decreases the uptake in the North Atlantic (Figure 10a). In both regions, inter-model differences are large (Figure 10b) and underline the uncertainty of F_{ant}^{ns}. The long-term trends in F_{ant}^{ns} are superimposed by an interannual and decadal variability that is mainly located in the subpolar North Atlantic and in the Southern Ocean (Figure 10c) and not in the Pacific Ocean as for F_{nat}^{ns} (Figures 4b and 4d). The interannual and decadal variability is similar across the entire model ensemble (Figure 10d).

Regionally, F_{ant}^{ns} is substantially smaller than F_{nat}^{ns} underlining the relatively minor importance of anthropogenic non-steady state fluxes compared to natural non-steady state fluxes. In the Southern Ocean, a strong negative trend in F_{ant}^{ns} co-occurs in regions with strong positive trends in F_{nat}^{ns} (Figure 4a). This suggests that both signals are related to stronger upwelling of circumpolar deep waters in most of the Southern Ocean due to recent trends in climate as also discussed by Hauck, Gregor, et al. (2023), Hauck, Nissen, et al. (2023), and Lovenduski et al. (2008). This increased upwelling brings more old waters containing higher concentrations of C_{nat} to the surface, enhancing the outgassing of C_{nat}. At the same time this exposes more waters to the surface with low concentrations of C_{ant}, causing an increase in F_{ant}^{ns}. In the subpolar North Atlantic, the strong positive F_{ant}^{ns} has a large model uncertainty associated with it, with some GOBMs showing a negative trend in F_{ant}^{ns}, while others show no significant trend. An independent model-study with one ESM (Goris et al., 2015) showed that the climate signal in the North Atlantic subpolar gyre is driven by counteracting processes (the influence of reduced biology and reduced circulation strength on DIC) and that relatively small differences in these contributions can shift this signal from a reduced pCO_2 to an increased pCO_2. Yet, their study considered an ESM with a large climate-induced AMOC decline and hence less warming in the subpolar gyre region, whereas the influence of warming can be of first order for models with a small AMOC decline (Bellomo et al., 2021). For RECCAP2, the timescale with climate change is not yet long-enough to separate the climate change signal from the strong decadal variability in the subpolar gyre and hence to attribute causes.

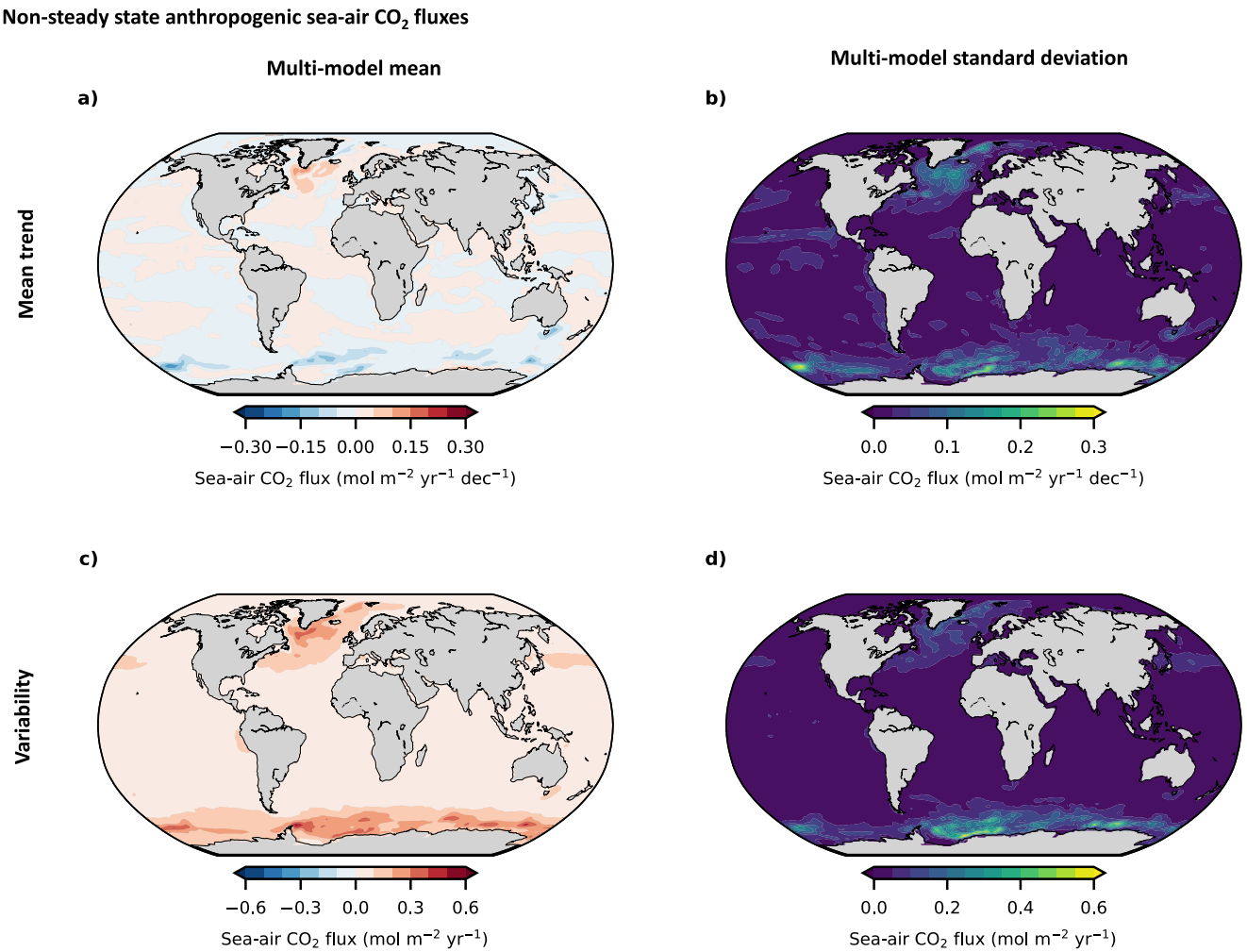

Figure 10. *Non-steady state anthropogenic sea-air CO_2 fluxes for eight global ocean biogeochemistry models*. Maps of the multi-model (a) mean of the linear trend and (b) standard deviation of the linear trend in anthropogenic non-steady state sea-air CO_2 flux averaged from 1980 to 2018, as well as maps of the multi-model (c) mean deviation of the inter-annual variability and (d) standard deviation of the inter-annual variability (linear trend is removed).

4. Discussion

4.1. Riverine and Sediment Fluxes

The inadequate or non-existing representation of the riverine and sediment fluxes in the GOBMs results in large uncertainties when quantifying F^{net} from these GOBMs. These uncertainties are partly caused by an inaccurate $F_{nat}^{riv\text{-}bur}$ caused by the inadequate boundary conditions (Section 3.1.1) and partly due to uncertainties of the observation-based $F_{obs}^{riv\text{-}bur}$ that were used instead of $F_{nat}^{riv\text{-}bur}$ when deriving F^{net} from the GOBM simulations (Section 2.2.4). While the here-considered global GOBM-estimates of F^{net} are unaffected by their simulated $F_{nat}^{riv\text{-}bur}$, the uncertainty of the observation-based estimate $F_{obs}^{riv\text{-}bur}$ affects global ocean F^{net} (Section 2.2.4). Regionally, however, both the simulated estimate of $F_{nat}^{riv\text{-}bur}$ and the observation-based estimate of $F_{obs}^{riv\text{-}bur}$ affect F^{net}.

Thus, an accurate observation-based estimate of the global $F_{obs}^{riv\text{-}bur}$ is necessary to estimate global F^{net} from GOBMs while the simulated $F_{nat}^{riv\text{-}bur}$ is still inaccurately low or non-existing. In addition, an accurate $F_{obs}^{riv\text{-}bur}$ also allows evaluation of simulated $F_{nat}^{riv\text{-}bur}$ by those GOBMs that aim at accurately representing $F_{net}^{riv\text{-}bur}$. Despite large efforts over the last years (Lacroix et al., 2020; Regnier et al., 2022; Resplandy et al., 2018), the most recent observation-based estimate of the global $F_{obs}^{riv\text{-}bur}$ of -0.65 ± 0.15 Pg C year^{-1} (Regnier et al., 2022) still has large uncertainties (1-sigma uncertainty of ~23%). These uncertainties are of similar magnitude as the simulated inter-model standard deviation of global F^{net} before accounting for $F_{obs}^{riv\text{-}bur}$ (1-sigma uncertainty of

± 0.24 Pg C year^{-1}) and thus represent most likely a major uncertainty for estimating F^{net} from GOBMs at the moment.

Regionally, the uncertainties in F^{net} caused by riverine and sediment fluxes are even larger than globally as $F_{nat}^{riv\text{-}bur}$ cannot be removed and no purely observation-based estimates for regional $F_{obs}^{riv\text{-}bur}$ exist. To remove uncertainties from the simulated $F_{nat}^{riv\text{-}bur}$, a better representation of riverine and sediment fluxes must be implemented. To derive a regional $F_{obs}^{riv\text{-}bur}$ for the RECCAP2 project, a spatially resolved simulated pattern of $F^{riv\text{-}bur}$ from a GOBM by Lacroix et al. (2020) was scaled up by a factor of 2.83 to match the global estimate of $F_{obs}^{riv\text{-}bur}$ by Regnier et al. (2022) (see Section 2.2.3). This upscaling introduces uncertainties as the pattern might change if the global flux is substantially larger. However, the upscaling was necessary because the globally integrated estimate of $F_{nat}^{riv\text{-}bur}$ by Lacroix et al. (2020) is around 3 times smaller than the global estimate of $F_{obs}^{riv\text{-}bur}$ by Regnier et al. (2022) and smaller than other previous estimates of global $F_{nat}^{riv\text{-}bur}$ (Aumont et al., 2001; Resplandy et al., 2018). One reason for the relatively small global estimate of $F_{obs}^{riv\text{-}bur}$ might be that Lacroix et al. (2020) quantify $F_{obs}^{riv\text{-}bur}$ as the difference between a simulation with observation-based riverine fluxes of carbon, alkalinity, and nutrients and a reference simulation in which carbon and nutrients were artificially added to each surface ocean grid cell, at the coast and in the open ocean, to equilibrate carbon and nutrient losses to the sediments. As a result, the signal of the subtraction of the artificial surface ocean carbon and nutrients input may override the riverine signal, especially in regions far away from river deltas such as the Southern Ocean and cause the globally too small $F_{obs}^{riv\text{-}bur}$. While this upscaling of a regional pattern is the only available estimate of regional $F_{obs}^{riv\text{-}bur}$, it remains hence highly uncertain.

Another uncertainty of regional $F_{obs}^{riv\text{-}bur}$ is stemming from the lability of the organic carbon that enters the ocean. Lacroix et al. (2020) assume relatively labile organic carbon from rivers, which results in a strong riverine-burial-induced carbon outgassing in ocean basins close to river deltas, such as in the Atlantic Ocean (0.27 Pg C year^{-1}), and a relatively weak riverine-burial-induced carbon outgassing in basins that are further away from river deltas, such as the Southern Ocean (0.04 Pg C year^{-1}). Contrarily, an older estimate by Aumont et al. (2001) assumes organic carbon to be more refractory and their results suggest a smaller $F_{obs}^{riv\text{-}bur}$ in the Atlantic Ocean and a larger $F_{obs}^{riv\text{-}bur}$ in the Southern Ocean. Less labile riverine organic matter can be transported far away from the river mouths in the Atlantic Ocean before it is remineralized and outgassed to the atmosphere. If only around a third of the estimated riverine-induced outgassing in the Atlantic Ocean by Lacroix et al. (2020) would instead occur in the Southern Ocean, F^{net} in the Atlantic Ocean would double. Hence, more refined estimates of the lability of organic matter and its effect on $F_{obs}^{riv\text{-}bur}$ are crucial to better constrain the net sea-air CO_2 flux and regional anthropogenic carbon sink estimates.

Additional uncertainties arise from fluxes of nutrients in rivers and into sediments, and changes of carbon, alkalinity, and nutrient fluxes over time. While riverine carbon and alkalinity fluxes directly affect surface ocean pCO_2 and hence F^{net}, riverine nutrient fluxes also affect surface ocean pCO_2 and F^{net} indirectly via changes in primary production and carbon export (Gao et al., 2023; Lacroix et al., 2020, 2021), especially in coastal oceans (Louchard et al., 2021) or the Arctic Ocean (Terhaar, Lauerwald, et al., 2021; Terhaar, Orr, Ethé, et al., 2019). In addition, the change in F^{net} over time due to changing riverine nutrient fluxes are neither simulated by the GOBMs nor represented in $F_{obs}^{riv\text{-}bur}$ from Regnier et al. (2022). The impact of changing riverine inputs on the ocean carbon sink depends in size and location on the prescribed riverine input and the model, as seen for CNRM-ESM2-1 (Séférian et al., 2019; Terhaar et al., 2022) and NorESM1-ME (Gao et al., 2023). The size and importance of both effects still remain relatively unknown and necessitate further research.

Overall, riverine-burial induced air-sea fluxes hence represent a major uncertainty when quantifying F^{net} from GOBMs and large efforts are needed to reduce this uncertainty. However, when quantifying the anthropogenic perturbation of the pre-industrial air-sea CO_2 flux (S^{OCEAN}) as in the Global Carbon Budget, this uncertainty does not affect the GOBM estimates (assuming $F^{riv\text{-}bur}$ is constant over time) but the estimate of S^{OCEAN} from pCO_2 products. As pCO_2 products directly estimate F^{net}, $F_{obs}^{riv\text{-}bur}$ has to be subtracted from F^{net} to estimate S^{OCEAN}.

4.2. Gap Between Trends in Observation-Based Estimates and GOBMs

Previous studies (Friedlingstein et al., 2022; Hauck et al., 2020) and our results here indicate that trends in the ocean sink since 2000 differ globally and regionally between GOBM estimates and pCO_2 products. Although these different trends suggest a divergence between GOBM estimates and pCO_2 products in recent years

(Figure 1a), this might be a misinterpretation. An overall increase in F^{net} by around ~20%, as suggested based on the here identified underestimation of the anthropogenic steady-state flux, would change this perception. The difference in F^{net} would not appear as a divergence of both estimates since 2000 but as a change from an underestimation of F^{net} by the pCO_2 products to an overestimation. Also, the observation-based estimate of $F^{riv\text{-}bur}$ is highly uncertain, and a more accurate estimate could lead to a shift in the baseline of the GOBMs as well. Nevertheless, the growth rates of F^{net} are different between GOBMs and pCO_2 products and uncertainties remain of how the ocean sink evolves.

In the Southern Ocean, the pCO_2 product estimate of the Southern Ocean carbon sink suggested that the variability before 2000 is mainly due to decadal variations (Gruber, Landschützer, & Lovenduski, 2019; Keppler & Landschützer, 2019; Landschützer et al., 2015; McKinley et al., 2017). Since 2000, the estimate of the pCO_2 products of the Southern Ocean carbon flux has been moving toward more uptake. While this ongoing increase in uptake based on the pCO_2 products of the Southern Ocean may just be a longer variability cycle, it could also indicate a disagreement on the trend of the ocean carbon sink between pCO_2-based and GOBM-based estimates for unknown reasons. Moreover, research (Gloege et al., 2021; Hauck, Nissen, et al., 2023) suggests that pCO_2 products of the Southern Ocean might overestimate the trend by 50%–130% and the amplitude of the decadal variability by 31% due to extrapolation of sparse observations with temporal aliasing.

The increasing gap in the Atlantic after 2000, appears to result from a smaller F^{net} trend in GOBMs than in pCO_2 products. Here, the observation-based estimate of $F^{riv\text{-}bur}$ is highly uncertain (see Section 4.1), and a different $F^{riv\text{-}bur}$ estimate could change the increasing gap to converging estimates. Nevertheless, there appears to be a smaller trend in GOBMs than in pCO_2 products, which can partly be explained by the negatively biased chemical uptake capacity of the GOBMs (Section 3.3.3.2). Related to this, Lebehot et al. (2019) showed for a suite of ESMs that the North Atlantic surface ocean fugacity of CO_2 increased at a significantly faster rate than observed and related this to substantial biases in alkalinity and its impact on the buffer capacity. The GOBMs also simulate a rather weak AMOC, whose influence on F_{ant}^{ss} increases with increasing atmospheric CO_2 (Section 3.3.3.1; Figure S6 in Supporting Information S1). Concurrently, the disagreement in Atlantic F^{net} trends between GOBMs and pCO_2 products is especially large in the subpolar North Atlantic (not shown, Pérez et al., 2023). The location of this disagreement is likely related to the here identified AMOC-biases as the influence of AMOC-biases on F_{ant}^{ss} is potentially highest in the subpolar gyre where subsurface waters low in C_{ant} outcrop. Furthermore, a study with ESMs has shown that AMOC-biases are strongly correlated to sea surface temperature biases in the North Atlantic (Wang et al., 2014). While we did not analyze sea surface temperature biases in the North Atlantic, Rodgers et al. (2023) found that the seasonal cycle of pCO_2 in the subpolar Atlantic is thermally driven in the GOBMs while that of the pCO_2-products is non-thermally driven. This might lead to the F^{net} of the GOBMs being more sensitive to warming than that based on pCO_2 products (Goris et al., 2018), which may contribute to the increasing gap between GOBMs and pCO_2-products with time. However, the magnitude of these contributions is unclear and remains to be identified.

4.3. Inter-Annual and Decadal Variability of the Sea-Air CO_2 Flux

The here-used GOBM simulations suggest that, for the time-period 1980–2018, the largest share of the inter-annual and decadal variability of F^{net} results from F_{nat}^{ns}, that is, the sea-air CO_2 flux of natural carbon due to climate variability and climate change. Globally, F_{nat}^{ns} is also an important flux component as it allows comparing the estimated ocean carbon sink from surface ocean pCO_2 products, which quantify F_{ant}^{ss}, F_{ant}^{ns}, F_{nat}^{ns}, and $F_{obs}^{riv\text{-}bur}$ (Friedlingstein et al., 2022) to observation-based estimates of the interior ocean change of C_{ant} (Gruber, Clement, et al., 2019), which quantifies only changes in F_{ant}^{ss}, and F_{ant}^{ns}.

Previous cumulative estimates of the global F_{nat}^{ns} from 1994 to 2007 are 5 ± 3 Pg C (Gruber, Clement, et al. (2019); based on inverse estimates of anthropogenic carbon fluxes, net surface air-sea CO_2 fluxes and river-derived CO_2-outgassing), 1.3 Pg C (Friedlingstein et al., 2022, based on GOBMs) and 1.6 ± 0.8 Pg C (Terhaar et al. (2022), based on ESM simulations). The GOBMs here estimate a cumulative F_{nat}^{ns} of 1.6 ± 0.8 Pg C over the same period, which is similar to both previous model-based estimates, although the ESM-based estimate accounts only for the effect of climate change and externally forced variability (volcanoes, variability in atmospheric CO_2) and not for the unforced variability of the climate system (e.g., winds, atmospheric temperature etc.).

Regionally, the variability of the sea-air CO_2 flux is similar between GOBMs and pCO_2 products in the Pacific Ocean, where most of the inter-annual variability is located, and differs in the Southern Ocean, where pCO_2

products suggest a strong decadal variability before 2000 and a quasi-linear trend after 2000 (Gloege et al., 2021; Gruber, Landschützer, & Lovenduski, 2019; Landschützer et al., 2015). However, the sparse observations in the Southern Ocean pose a challenge for the observation-based estimates, especially in the sea ice area (Hauck, Gregor, et al., 2023; Hauck, Nissen, et al., 2023). For example, Gloege et al. (2021) showed that the MPI-SOM-FFN method used by one of these methods (Landschützer et al., 2015) may overestimate the decadal variability in the Southern Ocean by 30%. Potential reasons for biases in the GOBMs might be uncertainties in the atmospheric reanalysis data, non-representation of freshwater fluxes, or a too low internal ocean variability in the GOBMs, causing too little variability in the upwelling of circumpolar deep water or variability in the extent of Antarctic sea ice. It remains an open question how strong the decadal variability of the ocean carbon sink in the Southern Ocean is and how it is driven.

In comparison to F_{nat}^{ns}, the largest F_{ant}^{ns} are simulated in the subpolar North Atlantic with yet unidentified drivers and in the Southern Ocean where sea ice retreats with global warming and westerly winds strengthen and shift southwards (Purich et al., 2016). The strengthening of F_{ant}^{ns} in the Southern Ocean could be explained by more upwelling of old water with low C_{ant} content (Le Quéré et al., 2007) allowing for more C_{ant} uptake or additional free ocean surface due to climate change, which can take up more C_{ant}. Both processes would lead to partial compensation by F_{nat}^{ns} fluxes (Hauck, Gregor, et al., 2023; Hauck, Nissen, et al., 2023; Lovenduski et al., 2008), with either more natural carbon being upwelled to the surface or more C_{nat} being released with reduced ice cover.

4.4. Comparison to Previous Evaluations of GOBMs

Previous studies have assessed GOBMs and their fidelity to simulate the ocean carbon sink globally and regionally when forced with atmospheric reanalysis (e.g., Fay & McKinley, 2021; Hauck et al., 2020). Hauck et al. (2020) found that GOBMs on average overestimate the observed pCO_2 from SOCAT (Bakker et al., 2016), which suggests an underestimation of the ocean carbon uptake by GOBMs. This is consistent with our assessment that suggests an underestimation of the simulated ocean carbon by GOBMs sink primarily because of circulation biases. The partially late-starting date and biases in the chemical uptake capacity in models also tend to enhance this underestimation. Fay and McKinley (2021) tested how well GOBMs resemble the pCO_2 products flux estimates regionally, thereby repeating an analysis from the RECCAP1 project by Séférian et al. (2014). By selecting the GOBMs that perform best, they suggest that the simulated global ocean carbon sink is smaller than previously estimated, opposite to what this study and Hauck et al. (2020) suggest. Several assumptions are made by Fay and McKinley (2021), such as the application of the local riverine adjustment by Lacroix et al. (2020), not accounting for each model's simulated regional $F_{nat}^{riv\text{-}bur}$, and that an area-weighted repartitioning $F^{drift+bias}$ over the entire ocean surface is valid. However, the local riverine adjustments come with large uncertainties (Section 4.1) and our analysis suggests that $F^{drift+bias}$ and $F_{nat}^{riv\text{-}bur}$ are not evenly distributed. These adjustments affect the regional F^{net} and don't allow for robust simulated estimates of the regional F^{net}. Therefore, constraining the global F^{net} with regional F^{net} appears to be prone to large uncertainties and we recommend rather using underlying physical and biogeochemical processes for such constraints.

5. Conclusions and Recommendations

Our analysis of GOBMs helps to explain inter-model differences and differences between estimates of the ocean carbon sink from pCO_2-products and ocean biogeochemistry models (DeVries et al., 2023; Friedlingstein et al., 2022). These differences can be divided into (a) differences in the simulation setups, that is, starting year and model spin-up, (b) dynamical differences, that is, model physics and biogeochemistry, and (c) differences in boundary fluxes across the land-sea and sea-sediment interfaces.

The differences in the simulation setups can be resolved relatively easily by (a) pre-defining a commonly used pre-industrial state before the increase of anthropogenic CO_2 (year and associated CO_2 mixing ratio) that is used by all participating GOBMs and (b) increasing the spin-up period to reduce the uncertainty of the simulated F^{net} in relation to model drift. This also allows to pinpoint weaknesses of the GOBMs and relationships across the GOBMs which are more apparent in steady-state.

For the preindustrial state, our recommendation to start all simulations at a common date before the industrialization and the associated atmospheric CO_2 increase would for most GOBMs require additional simulation years. For example, starting the simulations in 1765 where atmospheric CO_2 levels started to increase due to changes in land use (Khatiwala et al., 2009; see also Section 2.6) would necessitates to perform up to 85 more years per

simulation. Yet, the cost of running GOBMs in hindcast mode is much smaller than the cost of fully-coupled ESMs and computational constraints should thus not represent a major bottleneck. The additional simulation years may remove a global bias that is at least 0.04–0.06 Pg C year^{-1} in simulations that only started in 1850 (underestimation of the sink) but possibly 40% larger.

In model estimates of the global ocean carbon sink, the bias due to a too short spin-up is often already accounted for through subtraction of the air-sea carbon flux of the control simulation with constant atmospheric CO_2 and climatological forcing (F_{nat}^{ss}) and hence does not affect the global estimate of the Global Carbon Budget (Friedlingstein et al., 2022; Hauck et al., 2020). However, a too short spin-up does impact regional flux estimates, particularly in the Southern Ocean and there is no easy option to account for these regional spin-up related biases. Moreover, these biases also influence the surface ocean carbonate chemistry. Such spin-up related biases in the surface ocean carbonate chemistry can influence sea-air CO_2 fluxes directly and also limit the identification of ensemble wide biases via emergent constraints. A long spin-up may, however, lead to the emergence of new biases, if the models' own steady state differs strongly from the real ocean (Séférian et al., 2016). While a longer spin-up increases the computational costs, it provides a relatively simple way to reduce the uncertainty of the simulated F^{net} in relation to model drift and allows to pinpoint weaknesses of the GOBMs which are more apparent in steady-state. This paves the way for more complex adjustments related to the models' physics, biology, and carbonate chemistry.

Improving the dynamical representation of the ocean circulation and biogeochemistry is more difficult. However, two ESM-derived relationships between the anthropogenic carbon flux into the ocean and key parameters of associated model dynamics (AMOC, Southern Ocean inter-frontal sea surface salinity) provide robust relationships to adjust simulated anthropogenic carbon fluxes for these two key processes until the representation of these processes is improved in the models. Our results show that the considered GOBMs have especially large offsets in the AMOC (3.1 ± 5.2 Sv) and slightly overestimate the inter-frontal sea surface salinity in the Southern Ocean (0.03 ± 0.13), yet the constraint changed the original estimate of F_{ant}^{ss} in the Atlantic and Southern Oceans by less than 10% here. Both emerging relationships would likely have been stronger and helped to reduce uncertainties more if all simulations had used the same starting dates and pre-industrial pCO_2. As opposed to biases in the ocean circulation, biases in the ocean biogeochemistry could not be directly linked to sea-air CO_2 fluxes. Our recommendations for model setup will likely improve the robustness of these relationships and allow us to infer the influence of ocean circulation and biogeochemistry biases on anthropogenic carbon fluxes more clearly. In the short term, these constraints can be applied to account for model biases in circulation when estimating the ocean carbon sink from model ensembles, such as in the Global Carbon Budget (Friedlingstein et al., 2022). Though the emergent constraints have changed the original estimate only slightly here, other model ensembles might have larger biases and changes in F_{ant}^{ss} might hence be larger. In the long-term, we recommend more complex adjustments within the setups of the GOBMs to reduce these circulation biases.

The relatively poor representation of riverine and burial fluxes introduces another uncertainty to the simulated sea-air CO_2 fluxes. Simulated sea-air CO_2 fluxes caused by riverine and burial fluxes do not or poorly represent the observation-based estimate of this flux (Regnier et al., 2022), such that it remains challenging to compare the modeled estimates to the observation-based estimates of the ocean carbon sink. Although the representation of these fluxes and the resulting sea-air CO_2 fluxes do not directly influence the GOBM-based global ocean carbon sink estimated in the Global Carbon Budget (Friedlingstein et al., 2022), they affect the quantification of F^{net} and make a model quantification of natural sea-air CO_2 fluxes almost impossible due to their regionally large size and introduce large uncertainties for the estimation of regional net sea-air CO_2 fluxes. Until these sea-air CO_2 fluxes caused by riverine and burial fluxes are better simulated, an observation-based estimate of the pre-industrial sea-air CO_2 flux from riverine carbon, alkalinity, and nutrient input and its large uncertainty has to be added to the simulated flux by GOBMs to estimate F^{net} or has to be subtracted from the pCO_2 products to be able to compare these estimates the global carbon sink. While improvements in the global estimate of these pre-industrial sea-air CO_2 fluxes from riverine carbon and nutrient input have been recently made (e.g., Gao et al., 2023; Lacroix et al., 2020), the regional distribution and temporal variability of these fluxes remains highly uncertain and renders a comparison between simulated and observation-based estimates of the ocean carbon sink complicated. Improving the representation of these fluxes and their underlying processes in GOBMs and observation-based estimates is thus of importance to better understand the regional ocean carbon sinks.

The work here contributes to understanding the apparent gap between the growth rates of the carbon sink in model-based and pCO_2 product estimates. A number of different factors (a late starting date, circulation biases, biogeochemical biases, biases in C_{ant} storage) suggest that the GOBMs underestimate the ocean carbon sink on average and that this underestimation explains the apparent difference between long-term averaged carbon sink estimates of GOBMs and pCO_2 products. If the global ocean carbon sink estimate from GOBMs was on average higher, the different trends since 2000 in the GOBM estimate and pCO_2 products would not lead to a divergence of both estimates, but to a crossing from a weaker estimate from pCO_2 products to a stronger estimate from pCO_2 products. While the extrapolation of sparse observations might lead to an overestimation of trends in the pCO_2 products (especially in the Southern Ocean), the differences in the long-term mean carbon sink and in growth rates between GOBMs and pCO_2 products still need further research.

Overall, this model evaluation has helped to give recommendations for the setup not only of RECCAP2-simulations but also of other simulations and provides possible explanations for the offset between estimates of the mean ocean carbon sink. In the short term, the most important steps would be to start simulations at a common date before the industrialization and the associated atmospheric CO_2 increase and increase the spin-up time to bring the pre-industrial simulations as close as possible to a steady state and to make key output metrics relating to ocean circulation, biogeochemistry and the land-ocean interface available. In the long-term, a better representation of simulated riverine and burial boundary fluxes and of ocean circulation and biogeochemistry is of importance. Possible avenues to achieve a better representation of ocean dynamics are, for example, simulations with different atmospheric reanalysis sets to quantify the influence of the prescribed atmospheric boundary conditions as well as testing the influence of higher resolution for the GOBMs.

Conflict of Interest

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

Data Availability Statement

All of the data in this manuscript from GOBMs, pCO_2 products, and pCO_2/alkalinity are available in a public repository (Müller, 2023).

The Bern3D-LPX model output is available in a public repository (Terhaar, 2023).

The GOBMs are described in the following references: CESM-ETHZ (Yang & Gruber, 2016), CNRM-ESM2-1 (Séférian et al., 2019), EC-Earth3 (Döscher et al., 2022), FESOM REcoM LR (Hauck et al., 2020), MRI-ESM2-1 (Urakawa et al., 2020), NorESM-OC1.2 (Schwinger et al., 2016), ORCA025-GEOMAR (Physics are described in Madec et al. (2017), and biogeochemistry in Chien et al. (2022)), and ORCA1-LIM3-PISCES (Aumont et al., 2015).

The pCO_2 products are described in the following references: CMEMS-LSCE-FFNN (Chau et al., 2022), CSIR-ML6 (Gregor et al., 2019), OceanSODA-ETHZ (Gregor & Gruber, 2021), JMA-MLR (Iida et al., 2021), MPI-SOMFFN (Landschützer et al., 2014), JENA-MLS (Rödenbeck et al., 2013), UOEX_Wat20 (Watson et al., 2020), and NIES-FNN (Zeng et al., 2014).

The pCO_2/alkalinity products are described in the following references: CMEMS-LSCE-FFNN (Chau et al., 2022), OceanSODA-ETHZ (Gregor & Gruber, 2021), and JMA-MLR (Iida et al., 2021).

The Bern3D-LPX Earth System Model is described in the following references: Lienert and Joos (2018) and Roth et al. (2014).

The GLODAPv2 data set is described in Lauvset et al. (2016) and is publicly available (Olsen et al., 2017).

Acknowledgments
The authors thank one anonymous reviewer and Rik Wanninkhof for their helpful and constructive comments when reviewing the manuscript, they helped to improve the manuscript substantially. J. Terhaar acknowledges funding from the Woods Hole Oceanographic Institution Postdoctoral Scholar Program, and the Swiss National Science Foundation under Grants PZ00P2_209044 and 200020_200511. He thanks the science team and the crew of the R/V Sikuliaq for providing an excellent work environment for advancing this work during the cruise in the Arctic Ocean in November 2022. N. Goris acknowledges funding from the Norwegian Research Council through the project COLUMBIA (Grant 275268) and thanks Jean Negrel for his assistance with the preparation of RECCAP2-model and data products for her analysis as well as NORCE for providing internal support to work on RECCAP2. J. Terhaar, J.D. Müller, and N. Gruber acknowledge funding from the European Union's Horizon 2020 research and innovation program under grant agreement no. 821003 (project 4C, Climate–Carbon Interactions in the Current Century). F.F. Pérez was supported by the BOCATS2 (PID2019-104279GB-C21) project funded by MCIN/AEI/10.13039/501100011033 and contributed to WATER:iOS CSIC PTI. N. Gruber acknowledges further support from the European Union's Horizon 2020 research and innovation programme under grant agreement no. 821001 (SO-CHIC). T. DeVries acknowledges support from NSF Grant OCE-1948955. J. Hauck and R. Seferian were supported by the European Union's Horizon 2020 research and innovation programme under grant agreement No. 820989 (project COMFORT). J. Hauck acknowledges funding from the Initiative and Networking Fund of the Helmholtz Association (Helmholtz Young Investigator Group Marine Carbon and Ecosystem Feedbacks in the Earth System [MarESys], Grant VH-NG-1301) and from ERC-2022-STG OceanPeak (Grant 101077209). R. Seferian thanks the ESM2025 project under the grant agreement No. 101003536. The work reflects only the authors' views; the European Commission and their executive agency are not responsible for any use that may be made of the information the work contains.

References

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

Aumont, O., Orr, J. C., Monfray, P., Ludwig, W., Amiotte-Suchet, P., & Probst, J.-L. (2001). Riverine-driven interhemispheric transport of carbon. *Global Biogeochemical Cycles*, *15*(2), 393–405. https://doi.org/10.1029/1999GB001238

Bakker, D. C. E., Pfeil, B., Landa, C. S., Metzl, N., O'Brien, K. M., Olsen, A., et al. (2016). A multi-decade record of high-quality fCO_2 data in version 3 of the Surface Ocean CO_2 Atlas (SOCAT). *Earth System Science Data*, *8*(2), 383–413. https://doi.org/10.5194/essd-8-383-2016

Bellomo, K., Angeloni, M., Corti, S., & von Hardenberg, J. (2021). Future climate change shaped by inter-model differences in Atlantic meridional overturning circulation response. *Nature Communications*, *12*(1), 3659. https://doi.org/10.1038/s41467-021-24015-w

Boé, J., Hall, A., & Qu, X. (2009). September sea-ice cover in the Arctic Ocean projected to vanish by 2100. *Nature Geoscience*, *2*(5), 341–343. https://doi.org/10.1038/ngeo467

Bourgeois, T., Goris, N., Schwinger, J., & Tjiputra, J. F. (2022). Stratification constrains future heat and carbon uptake in the Southern Ocean between 30°S and 55°S. *Nature Communications*, *13*(1), 340. https://doi.org/10.1038/s41467-022-27979-5

Broecker, W. S., Takahashi, T., Simpson, H. J., & Peng, T.-H. (1979). Fate of fossil fuel carbon dioxide and the global carbon budget. *Science*, *206*(4417), 409–418. https://doi.org/10.1126/science.206.4417.409

Bronselaer, B., Winton, M., Russell, J., Sabine, C. L., & Khatiwala, S. (2017). Agreement of CMIP5 simulated and observed ocean anthropogenic CO_2 uptake. *Geophysical Research Letters*, *44*(24), 12298–12305. https://doi.org/10.1002/2017GL074435

Caldeira, K., & Duffy, P. B. (2000). The role of the Southern Ocean in uptake and storage of anthropogenic carbon dioxide. *Science*, *287*(5453), 620–622. https://doi.org/10.1126/science.287.5453.620

Chau, T. T. T., Gehlen, M., & Chevallier, F. (2022). A seamless ensemble-based reconstruction of surface ocean pCO_2 and air–sea CO_2 fluxes over the global coastal and open oceans. *Biogeosciences*, *19*(4), 1087–1109. https://doi.org/10.5194/bg-19-1087-2022

Chien, C.-T., Durgadoo, J. V., Ehlert, D., Frenger, I., Keller, D. P., Koeve, W., et al. (2022). FOCI-MOPS v1 - Integration of marine biogeochemistry within the Flexible Ocean and Climate Infrastructure version 1 (FOCI 1) Earth System Model. *Geoscientific Model Development*, *15*(15), 5987–6024. https://doi.org/10.5194/gmd-15-5987-2022

Clement, D., & Gruber, N. (2018). The eMLR(C*) method to determine decadal changes in the global ocean storage of anthropogenic CO_2. *Global Biogeochemical Cycles*, *32*(4), 654–679. https://doi.org/10.1002/2017GB005819

Cox, P. M., Pearson, D., Booth, B. B., Friedlingstein, P., Huntingford, C., Jones, C. D., & Luke, C. M. (2013). Sensitivity of tropical carbon to climate change constrained by carbon dioxide variability. *Nature*, *494*(7437), 341–344. https://doi.org/10.1038/nature11882

Davila, X., Gebbie, G., Brakstad, A., Lauvset, S. K., McDonagh, E. L., Schwinger, J., & Olsen, A. (2022). How is the ocean anthropogenic carbon reservoir filled? *Global Biogeochemical Cycles*, *36*(5), e2021GB007055. https://doi.org/10.1029/2021GB007055

DeVries, T. (2014). The oceanic anthropogenic CO_2 sink: Storage, air-sea fluxes, and transports over the industrial era. *Global Biogeochemical Cycles*, *28*(7), 631–647. https://doi.org/10.1002/2013GB004739

DeVries, T. (2022). Atmospheric CO_2 and sea surface temperature variability cannot explain recent decadal variability of the ocean CO_2 sink. *Geophysical Research Letters*, *49*(7), e2021GL096018. https://doi.org/10.1029/2021GL096018

DeVries, T., Yamamoto, K., Wanninkhof, R., Gruber, N., Hauck, J., Müller, J. D., et al. (2023). Magnitude, trends, and variability of the global ocean carbon sink from 1985-2018. *Global Biogeochemical Cycles*, *37*(10), e2023GB007780. https://doi.org/10.1029/2023GB007780

Dickson, A. G., Sabine, C. L., & Christian, J. R. (2007). *Guide to best practices for ocean CO_2 measurements*. North Pacific Marine Science Organization.

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

Eyring, V., Cox, P. M., Flato, G. M., Gleckler, P. J., Abramowitz, G., Caldwell, P., et al. (2019). Taking climate model evaluation to the next level. *Nature Climate Change*, *9*(2), 102–110. https://doi.org/10.1038/s41558-018-0355-y

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

Fay, A. R., & McKinley, G. A. (2021). Observed regional fluxes to constrain modeled estimates of the ocean carbon sink. *Geophysical Research Letters*, *48*(20), e2021GL095325. https://doi.org/10.1029/2021GL095325

Feely, R. A., Wanninkhof, R., Takahashi, T., & Tans, P. (1999). Influence of El Niño on the equatorial Pacific contribution to atmospheric CO_2 accumulation. *Nature*, *398*(6728), 597–601. https://doi.org/10.1038/19273

Fennel, K., Mattern, J. P., Doney, S. C., Bopp, L., Moore, A. M., Wang, B., & Yu, L. (2022). Ocean biogeochemical modelling. *Nature Reviews Methods Primers*, *2*(1), 76. https://doi.org/10.1038/s43586-022-00154-2

Frajka-Williams, E., Moat, B., Smeed, D., Rayner, D., Johns, W., Baringer, M., et al. (2021). *Atlantic meridional overturning circulation observed by the RAPID-MOCHA-WBTS (RAPID-Meridional Overturning Circulation and Heatflux Array-Western Boundary Time Series) array at 26N from 2004 to 2020 (v2020.1)*. NERC EDS British Oceanographic Data Centre NOC. https://doi.org/10.5285/cc1e34b3-3385-662b-e053-6c86abc03444

Friedlingstein, P., O'Sullivan, M., Jones, M. W., Andrew, R. M., Gregor, L., Hauck, J., et al. (2022). Global carbon budget 2022. *Earth System Science Data*, *14*(11), 4811–4900. https://doi.org/10.5194/essd-14-4811-2022

Friedlingstein, P., O'Sullivan, M., Jones, M. W., Andrew, R. M., Hauck, J., Olsen, A., et al. (2020). Global carbon budget 2020. *Earth System Science Data*, *12*(4), 3269–3340. https://doi.org/10.5194/essd-12-3269-2020

Fröb, F., Olsen, A., Våge, K., Moore, G. W. K., Yashayaev, I., Jeansson, E., & Rajasakaren, B. (2016). Irminger Sea deep convection injects oxygen and anthropogenic carbon to the ocean interior. *Nature Communications*, *7*(1), 13244. https://doi.org/10.1038/ncomms13244

Frölicher, T. L., Sarmiento, J. L., Paynter, D. J., Dunne, J. P., Krasting, J. P., & Winton, M. (2015). Dominance of the Southern Ocean in anthropogenic carbon and heat uptake in CMIP5 models. *Journal of Climate*, *28*(2), 862–886. https://doi.org/10.1175/JCLI-D-14-00117.1

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

Gao, S., Schwinger, J., Tjiputra, J., Bethke, I., Hartmann, J., Mayorga, E., & Heinze, C. (2023). Riverine impact on future projections of marine primary production and carbon uptake. *Biogeosciences*, *20*(1), 93–119. https://doi.org/10.5194/bg-20-93-2023

Gattuso, J.-P., Magnan, A. K., Bopp, L., Cheung, W. W. L., Duarte, C. M., Hinkel, J., et al. (2018). Ocean solutions to address climate change and its effects on marine ecosystems. *Frontiers in Marine Science*, *5*, 337. https://doi.org/10.3389/fmars.2018.00337

Gloege, 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

Gomez, F. A., Wanninkhof, R., Barbero, L., & Lee, S.-K. (2021). Increasing river alkalinity slows ocean acidification in the Northern Gulf of Mexico. *Geophysical Research Letters*, *48*(24), e2021GL096521. https://doi.org/10.1029/2021GL096521

Goris, N., Johannsen, K., & Tjiputra, J. (2023). The emergence of the Gulf Stream and interior western boundary as key regions to constrain the future North Atlantic carbon uptake. *Geoscientific Model Development*, *16*(8), 2095–2117. https://doi.org/10.5194/gmd-16-2095-2023

Goris, N., Tjiputra, J., Schwinger, J., & Heinze, C. (2015). Responses of carbon uptake and oceanic pCO_2 to climate change in the North Atlantic: A model study with the Bergen Earth System Model. *Global Biogeochemical Cycles*, *29*(10), 1567–1583. https://doi.org/10.1002/2015GB005109

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

Gregor, L., & Gruber, N. (2021). OceanSODA-ETHZ: A global gridded data set of the surface ocean carbonate system for seasonal to decadal studies of ocean acidification. *Earth System Science Data*, *13*(2), 777–808. https://doi.org/10.5194/essd-13-777-2021

Gregor, L., Lebehot, A. D., Kok, S., & Scheel Monteiro, P. M. (2019). A comparative assessment of the uncertainties of global surface ocean CO_2 estimates using a machine-learning ensemble (CSIR-ML6 version2019a) – Have we hit the wall? *Geoscientific Model Development*, *12*(12), 5113–5136. https://doi.org/10.5194/gmd-12-5113-2019

Griffies, S. M., Danabasoglu, G., Durack, P. J., Adcroft, A. J., Balaji, V., Böning, C. W., et al. (2016). OMIP contribution to CMIP6: Experimental and diagnostic protocol for the physical component of the Ocean Model Intercomparison Project. *Geoscientific Model Development*, *9*(9), 3231–3296. https://doi.org/10.5194/gmd-9-3231-2016

Gruber, N., Bakker, D. C. E., DeVries, T., Gregor, L., Hauck, J., Landschützer, P., et al. (2023). Trends and variability in the ocean carbon sink. *Nature Reviews Earth & Environment*, *4*(2), 119–134. https://doi.org/10.1038/s43017-022-00381-x

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

Gruber, N., Landschützer, P., & Lovenduski, N. S. (2019). The variable southern ocean carbon sink. *Annual Review of Marine Science. Annual Reviews*, *11*(1), 159–186. https://doi.org/10.1146/annurev-marine-121916-063407

Hall, A., Cox, P., Huntingford, C., & Klein, S. (2019). Progressing emergent constraints on future climate change. *Nature Climate Change*, *9*(4), 269–278. https://doi.org/10.1038/s41558-019-0436-6

Hauck, J., Gregor, L., Nissen, C., Patara, L., Hague, M., Mongwe, N. 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*, *381*(2249), 20220063. https://doi.org/10.1098/rsta.2022.0063

Hauck, J., Zeising, M., Le Quéré, C., Gruber, N., Bakker, D. C. E., Bopp, L., et al. (2020). Consistency and challenges in the ocean carbon sink estimate for the global carbon budget. *Frontiers in Marine Science*, *7*. https://doi.org/10.3389/fmars.2020.571720

Haugan, P. M., & Drange, H. (1996). Effects of CO_2 on the ocean environment. *Energy Conversion and Management*, *37*(6), 1019–1022. https://doi.org/10.1016/0196-8904(95)00292-8

Iida, Y., Takatani, Y., Kojima, A., & Ishii, M. (2021). Global trends of ocean CO_2 sink and ocean acidification: An observation-based reconstruction of surface ocean inorganic carbon variables. *Journal of Oceanography*, *77*(2), 323–358. https://doi.org/10.1007/s10872-020-00571-5

IPCC. (2021). Summary for policymakers. In V. Masson-Delmotte, P. Zhai, A. Pirani, S. L. Connors, C. Péan, S. Berger, et al. (Eds.), *Climate change 2021: The physical science basis. Contribution of Working Group I to the sixth assessment report of the Intergovernmental Panel on Climate Change*. Cambridge University Press.

Joos, F., Plattner, G.-K., Stocker, T. F., Marchal, O., & Schmittner, A. (1999). Global warming and marine carbon cycle feedbacks on future atmospheric CO_2. *Science*, *284*(5413), 464–467. https://doi.org/10.1126/science.284.5413.464

Joos, F., Roth, R., Fuglestvedt, J. S., Peters, G. P., Enting, I. G., von Bloh, W., et al. (2013). Carbon dioxide and climate impulse response functions for the computation of greenhouse gas metrics: A multi-model analysis. *Atmospheric Chemistry and Physics*, *13*(5), 2793–2825. https://doi.org/10.5194/acp-13-2793-2013

Keppler, L., & Landschützer, P. (2019). Regional wind variability modulates the southern ocean carbon sink. *Scientific Reports*, *9*(1), 7384. https://doi.org/10.1038/s41598-019-43826-y

Khatiwala, S., Primeau, F., & Hall, T. M. (2009). Reconstruction of the history of anthropogenic CO_2 concentrations in the ocean. *Nature*, *462*(7271), 346–349. https://doi.org/10.1038/nature08526

Khatiwala, S., Tanhua, T., Mikaloff Fletcher, S. E., Gerber, M., Doney, S. C., Graven, H. D., et al. (2013). Global ocean storage of anthropogenic carbon. *Biogeosciences*, *10*(4), 2169–2191. https://doi.org/10.5194/bg-10-2169-2013

Khatiwala, S., Visbeck, M., & Cane, M. A. (2005). Accelerated simulation of passive tracers in ocean circulation models. *Ocean Modelling*, *9*(1), 51–69. https://doi.org/10.1016/j.ocemod.2004.04.002

Kwiatkowski, L., Bopp, L., Aumont, O., Ciais, P., Cox, P. M., Laufkötter, C., et al. (2017). Emergent constraints on projections of declining primary production in the tropical oceans. *Nature Climate Change*, *7*(5), 355–358. https://doi.org/10.1038/nclimate3265

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

Lacroix, F., Ilyina, T., Mathis, M., Laruelle, G. G., & Regnier, P. (2021). Historical increases in land-derived nutrient inputs may alleviate effects of a changing physical climate on the oceanic carbon cycle. *Global Change Biology*, *27*(21), 5491–5513. https://doi.org/10.1111/gcb.15822

Landschützer, P., Gruber, N., Bakker, D. C. E., & Schuster, U. (2014). Recent variability of the global ocean carbon sink. *Global Biogeochemical Cycles*, *28*(9), 927–949. https://doi.org/10.1002/2014GB004853

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

Large, W. G., & Yeager, S. G. (2009). The global climatology of an interannually varying air–sea flux data set. *Climate Dynamics*, *33*(2), 341–364. https://doi.org/10.1007/s00382-008-0441-3

Lauvset, S. K., Key, R. M., Olsen, A., van Heuven, S., Velo, A., Lin, X., et al. (2016). A new global interior ocean mapped climatology: The 1°x1° GLODAP version 2. *Earth System Science Data*, *8*(2), 325–340. https://doi.org/10.5194/essd-8-325-2016

Lebehot, A. D., Halloran, P. R., Watson, A. J., McNeall, D., Ford, D. A., Landschützer, P., et al. (2019). Reconciling observation and model trends in North Atlantic surface CO_2. *Global Biogeochemical Cycles*, *33*(10), 1204–1222. https://doi.org/10.1029/2019GB006186

Le Quéré, C., Orr, J. C., Monfray, P., Aumont, O., & Madec, G. (2000). Interannual variability of the oceanic sink of CO_2 from 1979 through 1997. *Global Biogeochemical Cycles*, *14*(4), 1247–1265. https://doi.org/10.1029/1999GB900049

Le Quéré, C., Rödenbeck, C., Buitenhuis, E. T., Conway, T. J., Langenfelds, R., Gomez, A., et al. (2007). Saturation of the Southern Ocean CO_2 sink due to recent climate change. *Science*, *316*(5832), 1735–1738. https://doi.org/10.1126/science.1136188

Lienert, S., & Joos, F. (2018). A Bayesian ensemble data assimilation to constrain model parameters and land-use carbon emissions. *Biogeosciences*, *15*(9), 2909–2930. https://doi.org/10.5194/bg-15-2909-2018

Locarnini, R. A., Mishonov, A. V., Baranova, O. K., Boyer, T. P., Zweng, M. M., Garcia, H. E., et al. (2018). *World Ocean Atlas 2018, Volume 1: Temperature*. NOAA Atlas NESDIS 82. Retrieved from https://archimer.ifremer.fr/doc/00651/76338/

Louchard, D., Gruber, N., & Münnich, M. (2021). The impact of the Amazon on the biological pump and the air-sea CO_2 balance of the western tropical Atlantic. *Global Biogeochemical Cycles*, *35*(6), e2020GB006818. https://doi.org/10.1029/2020GB006818

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

Lueker, T. J., Dickson, A. G., & Keeling, C. D. (2000). Ocean pCO_2 calculated from dissolved inorganic carbon, alkalinity, and equations for K1 and K2: Validation based on laboratory measurements of CO_2 in gas and seawater at equilibrium. *Marine Chemistry*, *70*(1), 105–119. https://doi.org/10.1016/S0304-4203(00)00022-0

Madec, G., Bourdallé-Badie, R., Bouttier, P.-A., Bricaud, C., Bruciaferri, D., Calvert, D., et al. (2017). NEMO ocean engine. Retrieved from http://hdl.handle.net/2122/13309

Marshall, J., & Speer, K. (2012). Closure of the meridional overturning circulation through Southern Ocean upwelling. *Nature Geoscience*, *5*(3), 171–180. https://doi.org/10.1038/ngeo1391

Matear, R. J., Wong, C. S., & Xie, L. (2003). Can CFCs be used to determine anthropogenic CO_2? *Global Biogeochemical Cycles*, *17*(1), 1013. https://doi.org/10.1029/2001GB001415

Matsumoto, K., & Gruber, N. (2005). How accurate is the estimation of anthropogenic carbon in the ocean? An evaluation of the ΔC* method. *Global Biogeochemical Cycles*, *19*(3), GB3014. https://doi.org/10.1029/2004GB002397

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

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

McNeil, B. I., & Matear, R. J. (2013). The non-steady state oceanic CO_2 signal: Its importance, magnitude and a novel way to detect it. *Biogeosciences*, *10*(4), 2219–2228. https://doi.org/10.5194/bg-10-2219-2013

Mehrbach, C., Culberson, C. H., Hawley, J. E., & Pytkowicx, R. M. (1973). Measurement of the apparent dissociation constants of carbonic acid in seawater at atmospheric pressure. *Limnology and Oceanography*, *18*(6), 897–907. https://doi.org/10.4319/lo.1973.18.6.0897

Middelburg, J. J., Soetaert, K., & Hagens, M. (2020). Ocean alkalinity, buffering and biogeochemical processes. *Reviews of Geophysics*, *58*(3), e2019RG000681. https://doi.org/10.1029/2019RG000681

Mignone, B. K., Gnanadesikan, A., Sarmiento, J. L., & Slater, R. D. (2006). Central role of Southern Hemisphere winds and eddies in modulating the oceanic uptake of anthropogenic carbon. *Geophysical Research Letters*, *33*(1), L01604. https://doi.org/10.1029/2005GL024464

Mikaloff Fletcher, S. E., Gruber, N., Jacobson, A. R., Doney, S. C., Dutkiewicz, S., Gerber, M., et al. (2006). Inverse estimates of anthropogenic CO_2 uptake, transport, and storage by the ocean. *Global Biogeochemical Cycles*, *20*(2), GB2002. https://doi.org/10.1029/2005GB002530

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

Millero, F. J. (1995). Thermodynamics of the carbon dioxide system in the oceans. *Geochimica et Cosmochimica Acta*, *59*(4), 661–677. https://doi.org/10.1016/0016-7037(94)00354-O

Millero, F. J., Lee, K., & Roche, M. (1998). Distribution of alkalinity in the surface waters of the major oceans. *Marine Chemistry*, *60*(1), 111–130. https://doi.org/10.1016/S0304-4203(97)00084-4

Müller, J. D. (2023). RECCAP2-ocean data collection [Dataset]. Zenodo. https://doi.org/10.5281/zenodo.7990823

Müller, J. D., Gruber, N., Carter, B., Feely, R., Ishii, M., Lange, N., et al. (2023). Decadal trends in the oceanic storage of anthropogenic carbon from 1994 to 2014. *AGU Advances*, *4*, e2023AV000875. https://doi.org/10.1029/2023AV000875

Olsen, A., Key, R. M., Lauvset, S. K., Kozyr, A., Tanhua, T., Hoppema, M., et al. (2017). Global Ocean Data Analysis Project, Version 2 (GLODAPv2) (NCEI Accession 0162565) Version 2. [Dataset: GLODAPv2.2016, mapped]. NOAA National Centers for Environmental Information. https://doi.org/10.7289/v5kw5d97

Orr, J. C., & Epitalon, J.-M. (2015). Improved routines to model the ocean carbonate system: Mocsy 2.0. *Geoscientific Model Development*, *8*(3), 485–499. https://doi.org/10.5194/gmd-8-485-2015

Orr, J. C., Maier-Reimer, E., Mikolajewicz, U., Monfray, P., Sarmiento, J. L., Toggweiler, J. R., et al. (2001). Estimates of anthropogenic carbon uptake from four three-dimensional global ocean models. *Global Biogeochemical Cycles*, *15*(1), 43–60. https://doi.org/10.1029/2000GB001273

Orr, J. C., Najjar, R. G., Aumont, O., Bopp, L., Bullister, J. L., Danabasoglu, G., et al. (2017). Biogeochemical protocols and diagnostics for the CMIP6 Ocean Model Intercomparison Project (OMIP). *Geoscientific Model Development*, *10*(6), 2169–2199. https://doi.org/10.5194/gmd-10-2169-2017

Pérez, F. F., Mercier, H., Vázquez-Rodríguez, M., Lherminier, P., Velo, A., Pardo, P. C., et al. (2013). Atlantic Ocean CO_2 uptake reduced by weakening of the meridional overturning circulation. *Nature Geoscience*, *6*(2), 146–152. https://doi.org/10.1038/ngeo1680

Pérez, F. F., Perez, F. F., Becker, M., Goris, N., Gehlen, M., Lopez-Mozos, M., et al. (2023). An assessment of CO_2 storage and sea-air fluxes for the Atlantic Ocean and Mediterranean Sea between 1985 and 2018. ESS Open Archive. December 27, 2023. https://doi.org/10.22541/essoar.170256825.55098483/v2

Planchat, A., Kwiatkowski, L., Bopp, L., Torres, O., Christian, J. R., Butenschön, M., et al. (2023). The representation of alkalinity and the carbonate pump from CMIP5 to CMIP6 Earth System Models and implications for the carbon cycle. *Biogeosciences*, *20*(7), 1195–1257. https://doi.org/10.5194/bg-20-1195-2023

Poulter, B., Bastos, A., Canadell, J., Ciais, P., Gruber, N., Hauck, J., et al. (2022). Inventorying Earth's land and ocean greenhouse gases. *Eos*, *103*. https://doi.org/10.1029/2022EO179084

Purich, A., Cai, W., England, M. H., & Cowan, T. (2016). Evidence for link between modelled trends in Antarctic sea ice and underestimated westerly wind changes. *Nature Communications*, *7*(1), 10409. https://doi.org/10.1038/ncomms10409

Regnier, P., Friedlingstein, P., Ciais, P., Mackenzie, F. T., Gruber, N., Janssens, I. A., et al. (2013). Anthropogenic perturbation of the carbon fluxes from land to ocean. *Nature Geoscience*, *6*(8), 597–607. https://doi.org/10.1038/ngeo1830

Regnier, P., Resplandy, L., Najjar, R. G., & Ciais, P. (2022). The land-to-ocean loops of the global carbon cycle. *Nature*, *603*(7901), 401–410. https://doi.org/10.1038/s41586-021-04339-9

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

Revelle, R., & Suess, H. E. (1957). Carbon dioxide exchange between atmosphere and ocean and the question of an increase of atmospheric CO_2 during the past decades. *Tellus*, *9*(1), 18–27. https://doi.org/10.1111/j.2153-3490.1957.tb01849.x

Ridge, S. M., & McKinley, G. A. (2020). Advective controls on the North Atlantic anthropogenic carbon sink. *Global Biogeochemical Cycles*, *34*(7), e2019GB006457. https://doi.org/10.1029/2019GB006457

Rödenbeck, C., Keeling, R. F., Bakker, D. C. E., Metzl, N., Olsen, A., Sabine, C., & Heimann, M. (2013). Global surface-ocean pCO_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., 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

Roth, R., Ritz, S. P., & Joos, F. (2014). Burial-nutrient feedbacks amplify the sensitivity of atmospheric carbon dioxide to changes in organic matter remineralisation. *Earth System Dynamics*, *5*(2), 321–343. https://doi.org/10.5194/esd-5-321-2014

Sabine, C. L., Feely, R. A., Gruber, N., Key, R. M., Lee, K., Bullister, J. L., et al. (2004). The oceanic sink for anthropogenic CO_2. *Science*, *305*(5682), 367–371. https://doi.org/10.1126/science.1097403

Sallée, J.-B., Shuckburgh, E., Bruneau, N., Meijers, A. J. S., Bracegirdle, T. J., Wang, Z., & Roy, T. (2013). Assessment of Southern Ocean water mass circulation and characteristics in CMIP5 models: Historical bias and forcing response. *Journal of Geophysical Research: Oceans*, *118*(4), 1830–1844. https://doi.org/10.1002/jgrc.20135

Sarmiento, J. L., & Gruber, N. (2006). *Ocean biogeochemical dynamics*. Princeton University Press. https://doi.org/10.1515/9781400849079

Sarmiento, J. L., Orr, J. C., & Siegenthaler, U. (1992). A perturbation simulation of CO_2 uptake in an ocean general circulation model. *Journal of Geophysical Research*, *97*(C3), 3621–3645. https://doi.org/10.1029/91JC02849

Sarmiento, J. L., & Sundquist, E. T. (1992). Revised budget for the oceanic uptake of anthropogenic carbon dioxide. *Nature*, *356*(6370), 589–593. https://doi.org/10.1038/356589a0

Schwinger, J., Goris, N., Tjiputra, J. F., Kriest, I., Bentsen, M., Bethke, I., et al. (2016). Evaluation of NorESM-OC (versions 1 and 1.2), the ocean carbon-cycle stand-alone configuration of the Norwegian Earth System Model (NorESM1). *Geoscientific Model Development*, *9*(8), 2589–2622. https://doi.org/10.5194/gmd-9-2589-2016

Séférian, R., Gehlen, M., Bopp, L., Resplandy, L., Orr, J. C., Marti, O., et al. (2016). Inconsistent strategies to spin up models in CMIP5: Implications for ocean biogeochemical model performance assessment. *Geoscientific Model Development*, *9*(5), 1827–1851. https://doi.org/10.5194/gmd-9-1827-2016

Séférian, 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

Séférian, R., Ribes, A., & Bopp, L. (2014). Detecting the anthropogenic influences on recent changes in ocean carbon uptake. *Geophysical Research Letters*, *41*(16), 5968–5977. https://doi.org/10.1002/2014GL061223

Steinfeldt, R., Rhein, M., Bullister, J. L., & Tanhua, T. (2009). Inventory changes in anthropogenic carbon from 1997–2003 in the Atlantic Ocean between 20°S and 65°N. *Global Biogeochemical Cycles*, *23*(3), GB3010. https://doi.org/10.1029/2008GB003311

Stock, C. A., Dunne, J. P., Fan, S., Ginoux, P., John, J., Krasting, J. P., et al. (2020). Ocean biogeochemistry in GFDL's Earth System Model 4.1 and its response to increasing atmospheric CO_2. *Journal of Advances in Modeling Earth Systems*, *12*(10), e2019MS002043. https://doi.org/10.1029/2019MS002043

Takahashi, T., Broecker, W. S., & Bainbridge, a. E. (1981). The alkalinity and total carbon dioxide concentration in the world oceans. *Carbon Cycle Modelling, SCOPE*, *16*(3078), 271–286. Retrieved from http://data.ieda.org.cn/upload/fckeditor/SCOPE_16_1.5.07_Takahashi_271-286.pdf

Talley, L. D. (2013). Closure of the global overturning circulation through the Indian, Pacific, and Southern Oceans: Schematics and transports. *Oceanography*, *26*(1), 80–97. https://doi.org/10.5670/oceanog.2013.07

Tanhua, T., Hoppema, M., Jones, E. M., Stöven, T., Hauck, J., Dávila, M. G., et al. (2017). Temporal changes in ventilation and the carbonate system in the Atlantic sector of the Southern Ocean. *Deep-Sea Research Part II: Topical Studies in Oceanography*, *138*, 26–38. https://doi.org/10.1016/j.dsr2.2016.10.004

Terhaar, J. (2023). Ocean carbon uptake in simulations with Bern3D-LPX with different starting dates [Dataset]. SEANOE. https://doi.org/10.17882/95124

Terhaar, J., Frölicher, T. L., & Joos, F. (2022). Observation-constrained estimates of the global ocean carbon sink from Earth System Models. *Biogeosciences*, *19*(18), 4431–4457. https://doi.org/10.5194/bg-19-4431-2022

Terhaar, J., Frölicher, T., & Joos, F. (2021). Southern Ocean anthropogenic carbon sink constrained by sea surface salinity. *Science Advances*, *7*(18), 5964–5992. https://doi.org/10.1126/sciadv.abd5964

Terhaar, J., Frölicher, T. L., & Joos, F. (2023). Ocean acidification in emission-driven temperature stabilization scenarios: The role of TCRE and non-CO_2 greenhouse gases. *Environmental Research Letters*, *18*(2), 024033. https://doi.org/10.1088/1748-9326/acaf91

Terhaar, J., Kwiatkowski, L., & Bopp, L. (2020). Emergent constraint on Arctic Ocean acidification in the twenty-first century. *Nature*, *582*(7812), 379–383. https://doi.org/10.1038/s41586-020-2360-3

Terhaar, J., Lauerwald, R., Regnier, P., Gruber, N., & Bopp, L. (2021). Around one third of current Arctic Ocean primary production sustained by rivers and coastal erosion. *Nature Communications*, *12*(1), 169. https://doi.org/10.1038/s41467-020-20470-z

Terhaar, J., Orr, J. C., Ethé, C., Regnier, P., & Bopp, L. (2019). Simulated Arctic Ocean response to doubling of riverine carbon and nutrient delivery. *Global Biogeochemical Cycles*, *33*(8), 1048–1070. https://doi.org/10.1029/2019GB006200

Terhaar, J., Orr, J. C., Gehlen, M., Ethé, C., & Bopp, L. (2019). Model constraints on the anthropogenic carbon budget of the Arctic Ocean. *Biogeosciences*, *16*(11), 2343–2367. https://doi.org/10.5194/bg-16-2343-2019

Terhaar, J., Tanhua, T., Stöven, T., Orr, J. C., & Bopp, L. (2020). Evaluation of data-based estimates of anthropogenic carbon in the Arctic Ocean. *Journal of Geophysical Research: Oceans*, *125*(6), e2020JC016124. https://doi.org/10.1029/2020jc016124

Terhaar, J., Torres, O., Bourgeois, T., & Kwiatkowski, L. (2021). Arctic Ocean acidification over the 21st century co-driven by anthropogenic carbon increases and freshening in the CMIP6 model ensemble. *Biogeosciences*, *18*(6), 2221–2240. https://doi.org/10.5194/bg-18-2221-2021

Tsujino, H., Urakawa, S., Nakano, H., Small, R. J., Kim, W. M., Yeager, S. G., et al. (2018). JRA-55 based surface dataset for driving ocean–sea-ice models (JRA55-do). *Ocean Modelling*, *130*, 79–139. https://doi.org/10.1016/j.ocemod.2018.07.002

Uppström, L. R. (1974). The boron/chlorinity ratio of deep-sea water from the Pacific Ocean. *Deep-Sea Research and Oceanographic Abstracts*, *21*(2), 161–162. https://doi.org/10.1016/0011-7471(74)90074-6

Urakawa, L. S., Tsujino, H., Nakano, H., Sakamoto, K., Yamanaka, G., & Toyoda, T. (2020). The sensitivity of a depth-coordinate model to diapycnal mixing induced by practical implementations of the isopycnal tracer diffusion scheme. *Ocean Modelling*, *154*, 101693. https://doi.org/10.1016/j.ocemod.2020.101693

Vaittinada Ayar, P., Bopp, L., Christian, J. R., Ilyina, T., Krasting, J. P., Séférian, R., et al. (2022). Contrasting projections of the ENSO-driven CO_2 flux variability in the equatorial Pacific under high-warming scenario. *Earth System Dynamics*, *13*(3), 1097–1118. https://doi.org/10.5194/esd-13-1097-2022

Wang, Q., Danilov, S., Sidorenko, D., Timmermann, R., Wekerle, C., Wang, X., et al. (2014). The Finite Element Sea Ice-Ocean Model (FESOM) v.1.4: Formulation of an ocean general circulation model. *Geoscientific Model Development*, *7*(2), 663–693. https://doi.org/10.5194/gmd-7-663-2014

Wanninkhof, R., Park, G.-H., Takahashi, T., Sweeney, C., Feely, R., Nojiri, Y., et al. (2013). Global ocean carbon uptake: Magnitude, variability and trends. *Biogeosciences*, *10*(3), 1983–2000. https://doi.org/10.5194/bg-10-1983-2013

Watson, A. J., Schuster, U., Shutler, J. D., Holding, T., Ashton, I. G. C., Landschützer, P., et al. (2020). Revised estimates of ocean-atmosphere CO_2 flux are consistent with ocean carbon inventory. *Nature Communications*, *11*(1), 4422. https://doi.org/10.1038/s41467-020-18203-3

Waugh, D. W., Hall, T. M., McNeil, B. I., Key, R. M., & Matear, R. J. (2006). Anthropogenic CO_2 in the oceans estimated using transit time distributions. *Tellus B*, *58*(5), 376–389. https://doi.org/10.1111/j.1600-0889.2006.00222.x

Waugh, D. W., Primeau, F., DeVries, T., & Holzer, M. (2013). Recent changes in the ventilation of the Southern Oceans. *Science*, *339*(6119), 568–570. https://doi.org/10.1126/science.1225411

Weiss, R. F. (1974). Carbon dioxide in water and seawater: The solubility of a non-ideal gas. *Marine Chemistry*, *2*(3), 203–215. https://doi.org/10.1016/0304-4203(74)90015-2

Wright, R. M., Le Quéré, C., Buitenhuis, E., Pitois, S., & Gibbons, M. J. (2021). Role of jellyfish in the plankton ecosystem revealed using a global ocean biogeochemical model. *Biogeosciences*, *18*(4), 1291–1320. https://doi.org/10.5194/bg-18-1291-2021

Yang, S., & Gruber, N. (2016). The anthropogenic perturbation of the marine nitrogen cycle by atmospheric deposition: Nitrogen cycle feedbacks and the ^{15}N Haber-Bosch effect. *Global Biogeochemical Cycles*, *30*(10), 1418–1440. https://doi.org/10.1002/2016GB005421

Yasunaka, S., Manizza, M., Terhaar, J., Olsen, A., Yamaguchi, R., Landschützer, P., et al. (2023). An assessment of CO_2 uptake in the Arctic Ocean from 1985 to 2018. *Global Biogeochemical Cycles*, *37*(11), e2023GB007806. https://doi.org/10.1029/2023GB007806

Zeng, J., Nojiri, Y., Landschützer, P., Telszewski, M., & Nakaoka, S. (2014). A global surface ocean fCO_2 climatology based on a feed-forward neural network. *Journal of Atmospheric and Oceanic Technology*, *31*(8), 1838–1849. https://doi.org/10.1175/JTECH-D-13-00137.1

Zweng, M. M., Reagan, J. R., Seidov, D., Boyer, T. P., Locarnini, R. A., Garcia, H. E., et al. (2018). World Ocean Atlas 2018, Volume 2: Salinity. Retrieved from https://www.ncei.noaa.gov/sites/default/files/2020-04/woa18_vol2.pdf