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

Jens Terhaar¹, Nadine Goris², Jens Daniel Müller³, Timothy DeVries⁴, Nicolas Gruber³, Judith Hauck⁵, Fiz F. Pérez⁶, and Roland Séférian⁷

¹Woods Hole Oceanographic Institution
²NORCE
³ETH Zürich
⁴University of California Santa Barbara
⁵Alfred Wegener Institute Helmholtz Centre for Polar and Marine Research
⁶Instituto de Investigaciones Marinas (CSIC)
⁷CNRM (Météo-France/CNRS)

May 13, 2023

Abstract

The ocean is a major carbon sink and takes up 25-30% of the anthropogenically emitted CO₂. A state-of-the-art method to quantify this sink are global ocean biogeochemistry models (GOBMs) but their simulated CO₂ uptake differs between models and is systematically lower than estimates based on statistical methods using surface ocean pCO₂ 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 the (i) model set-up, such as the length of the spin-up, the starting date of the simulation, and carbon fluxes from rivers and into sediments, (ii) the ocean circulation, such as Atlantic Meridional Overturning Circulation and Southern Ocean mode and intermediate water formation, and (iii) the oceanic buffer capacity. Our analysis suggests that the late starting date and biases in the ocean circulation cause a too low anthropogenic CO₂ uptake across the GOBM ensemble. Surface ocean biogeochemistry biases might also cause simulated anthropogenic fluxes to be too low but the current set-up prevents a robust assessment. For simulations of the ocean carbon sink, we recommend in the short-term to (1) start simulations in 1765, when atmospheric CO₂ started to increase, (2) conduct a sufficiently long spin-up such that the GOBMs reach steady-state, and (3) 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.
Assessment of Global Ocean Biogeochemistry Models for Ocean Carbon Sink
Estimates in RECCAP2 and Recommendations for Future Studies

Jens Terhaar1,2,3,†, Nadine Goris4,†, Jens D. Müller5, Tim DeVries6,7, Nicolas Gruber5, Judith
Hauck8, Fiz F. Perez9, Roland Séférian10

1Department of Marine Chemistry and Geochemistry, Woods Hole Oceanographic Institution, Woods Hole, 02543, Massachusetts, USA
2Climate and Environmental Physics, Physics Institute, University of Bern, Switzerland
3Oeschger Centre for Climate Change Research, University of Bern, Switzerland
4NORCE Climate & Environment, Bjerknes Centre for Climate Research, Bergen, Norway
5Environmental Physics, Institute of Biogeochemistry and Pollutant Dynamics, ETH Zurich, Zürich, Switzerland
6Department of Geography, University of California, Santa Barbara, Santa Barbara, CA, USA
7Earth Research Institute, University of California, Santa Barbara, Santa Barbara, CA, USA
8Alfred-Wegener-Institut, Helmholtz-Zentrum für Polar- und Meeresforschung, Am Handelshafen 12, 27570 Bremerhaven, Germany
9Instituto de Investigaciones Marinas (IIM), CSIC, Vigo, Spain. (0000-0003-4836-8974)
10CNRM (Université de Toulouse, Météo-France, CNRS), Toulouse, France

Corresponding author: Jens Terhaar (jens.terhaar@unibe.ch)
†Nadine Goris and Jens Terhaar should be considered joint first authors.

Key Points:
• The simulated CO2-uptake by global ocean biogeochemistry models in RECCAP2 is systematically lower than observation-based estimates
• This underestimation is caused by the set-up of the RECCAP2-simulations as well as biases in surface chemistry and ocean circulation.
• Concrete steps forward are proposed to improve simulations of the ocean carbon sink by global ocean biogeochemistry models
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 $p$CO$_2$ 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 the (i) model set-up, such as the length of the spin-up, the starting date of the simulation, and carbon fluxes from rivers and into sediments, (ii) the ocean circulation, such as Atlantic Meridional Overturning Circulation and Southern Ocean mode and intermediate water formation, and (iii) the oceanic buffer capacity. Our analysis suggests that the 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 set-up prevents a robust assessment. For simulations of the ocean carbon sink, we recommend in the short-term to (1) start simulations in 1765, when atmospheric CO$_2$ started to increase, (2) conduct a sufficiently long spin-up such that the GOBMs reach steady-state, and (3) 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) with regards to their simulated CO$_2$ uptake from 1980 to 2018. We focus our analysis on the simulation set-up from the Global Carbon Budget initiative and the GOBMs that are used in the current version of the Global Carbon Budget. We find that the simulated CO$_2$ uptake by GOBMs is systematically lower than that of observation-based estimates and that the estimates differ substantially between GOBMs. We identify several reasons for this underestimation, relating to the set up of the simulations as well as the set up of the GOBMs themselves. For the set-up of the simulations, we find that not all GOBMs had the same starting year and the same initial stability, while the set up of the GOBMs themselves showed that the majority of GOBMs underestimate the large scale ocean circulation in the Atlantic and do not provide the necessary output for evaluation of their land-ocean interface. Based on our evaluation, we give recommendations for the set-up of follow up studies.

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; Terhaar et al., 2022), thereby reducing the growth of atmospheric CO$_2$ and slowing down global warming (IPCC, 2021). 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 CO₂ concentrations is known as anthropogenic carbon \( (C_{\text{ant}}; \text{Sarmiento et al., 1992}) \), while the DIC that existed prior to the start of the industrial revolution is called natural carbon \( (C_{\text{nat}}) \). 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 CO₂ fluxes \((\text{Joos et al., 1999; McNeil & Matear, 2013; Le Quéré et al., 2000})\), as well as anthropogenic changes in the riverine carbon fluxes \((\text{Regnier et al., 2013; Terhaar et al., 2022})\). At the global scale, the air-sea \( C_{\text{ant}} \) flux is controlled by the rate of \( C_{\text{ant}} \) transport from the surface ocean to the deep ocean, which depends on the concentration of \( C_{\text{ant}} \) in the surface ocean \((\text{Broecker et al., 1979})\) and the surface-to-deep water volume transport \((\text{Caldeira & Duffy, 2000; Mikaloff Fletcher et al., 2006; Orr et al., 2001; Sarmiento et al., 1992})\). In contrast, the air-sea flux of \( C_{\text{nat}} \) is primarily controlled by the interaction of surface heating or cooling affecting the solubility of CO₂ in seawater and transport and mixing, and biological processes of photosynthesis, respiration, and CaCO₃ production \((\text{Sarmiento & Gruber, 2006})\). While there is agreement on these drivers for \( C_{\text{ant}} \) and \( C_{\text{nat}} \) 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 \( \text{(RECCAP2)} \) \((\text{Poulter et al., 2022})\). 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 CO₂ \( (pCO₂ \text{ products}) \) and hindcast simulations from global and regional ocean biogeochemistry models \( \text{(GOBMs)} \) 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 \( pCO₂ \) products, the compiled database of RECCAP2 goes well beyond that used in the framework of the Global Carbon Budget \((\text{Friedlingstein et al., 2022})\). 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 CO₂ 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 \( \text{(RECCAP2)} \)” and in the 2020 and 2022 edition of the Global Carbon Budget \((\text{Friedlingstein et al., 2020, 2022})\) 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 \( pCO₂ \) products \((\text{Chau et al., 2022; Gregor et al., 2019; Gregor & Gruber, 2021; Iida et al., 2021; Landschützer et al., 2014; Rödenbeck et al., 2013; Watson et al., 2020; Zeng et al., 2014})\) and GOBMs forced with historic atmospheric reanalysis data \((\text{Hauck et al., 2020; Orr et al., 2001; Sarmiento et al., 1992; Sarmiento & Sundquist, 1992})\). The global ocean carbon sink estimates differ across the different methods and models with the multi-model mean simulated net oceanic carbon sink reported by the Global Carbon Budget being consistently less negative (lower uptake) than the mean estimate of the \( pCO₂ \)-products \((1990s: -1.91±0.25 \text{Pg C yr}^{-1} \text{ in models vs } -2.14±0.34 \text{Pg C yr}^{-1} \text{ for } pCO₂ \text{ products, 2000s: -2.05±0.27 } \text{Pg C yr}^{-1} \text{ vs } -2.34±0.21 \text{Pg C yr}^{-1}, \text{ and 2010s: -2.42±0.29 } \text{Pg C yr}^{-1} \text{ vs } -3.02±0.22 \text{ Pg C yr}^{-1} \text{; Friedlingstein et al., 2022})\). The difference between the models and \( pCO₂ \) products in the 2010s is around half as large as the annual CO₂ 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 the mode, intermediate, and deep-water formation in the North Atlantic (Goris et al., 2018; Terhaar et al., 2022) and Southern Ocean (Bourgeois et al., 2022; Fu et al., 2022; Terhaar, Frölicher, et al., 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 CO₂ uptake (Terhaar et al., 2022; Vaittinada Ayar et al., 2022). Other reasons for the above-mentioned differences are related to the set-up of the model simulations. For example, the starting date of model simulations is often several decades delayed relative to the onset of the anthropogenic CO₂ increase in the atmosphere around 1765 (Bronsema et al., 2017; Terhaar, Orr, Gehlen, et al., 2019), leading to too low ocean carbon uptake and storage. Associated with the set-up 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 pCO₂ 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 simulate their cycling through biogeochemical processes (primary production, grazing, remineralization, etc.) (Fennel et al., 2022). The air-sea CO₂ flux in these models is based on the simulated ocean carbon dynamics and the prescribed atmospheric CO₂ mixing ratio. In this study, we analyzed the following 8 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 CO₂ 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 also explain the reasons for its exclusion there, we exclude it from all GOBM results in terms of multi-model mean and standard deviation.

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 an overview table in DeVries et al. (in review). 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 simulations, the salinity restoring timescale was two 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 timescale in the Southern Ocean better captures sea-ice related fluxes that are not well represented in the atmospheric forcing fields. This change in the salinity restoring led to an improvement of the modeled overturning circulation, not only in the Southern Ocean, but also in the North Atlantic, where the previously very weak Atlantic Meridional Overturning Circulation (AMOC) increased from 3.5 Sv to 14.4 Sv (years 2005 to 2018).

2.2 Sea-air CO₂ flux

2.2.1 Sign convention

Throughout this study, the CO₂ flux between the atmosphere and ocean is defined as a sea-to-air flux, thus with a negative flux indicating an uptake of CO₂ 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 CO₂ flux

We followed the RECCAP2-ocean protocol and divided the total sea-air CO₂ flux ($F_{\text{total}}$) into five components. Specifically, the anthropogenic sea-air CO₂ flux from increasing atmospheric CO₂ in the atmosphere ($F_{\text{ant}}$) was divided into a steady-state component $F_{\text{ant ss}}$ representing the anthropogenic uptake flux in the absence of climate change and climate variability, and into a non-steady state component $F_{\text{ant ns}}$ reflecting the effect of climate change and climate variability on $F_{\text{ant}}$. Like $F_{\text{ant}}$, the natural sea-air flux of CO₂ under pre-industrial atmospheric CO₂ ($F_{\text{nat}}$) was divided into $F_{\text{nat}}$ under a constant climate (steady-state $F_{\text{nat ss}}$ or short $F_{\text{nat ns}}$), and the modulation of the $F_{\text{nat}}$ due to climate variability and climate change (non-steady
The fifth flux component is the sea-air CO\textsubscript{2} flux due to the difference between the input of carbon and alkalinity across the land-sea interfaces from rivers and coastal erosion and the burial of carbon and alkalinity components in sediments ($F_{\text{nat riv-bur}}$). 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 (RECCA P2)” consider $F_{\text{nat riv-bur}}$ to be an integral part of $F_{\text{nat ss}}$. We kept them separated to the degree that this is possible in order to analyze all flux components individually.

The total flux across the sea-air interface ($F_{\text{total}}$) can thus be written as:

$$F_{\text{total}} = F_{\text{ant ss}} + F_{\text{ant ns}} + F_{\text{nat ss}} + F_{\text{nat ns}} + F_{\text{nat riv-bur}}$$  \hspace{1cm} (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.

2.2.3 RECCAP2 simulations and their relation to CO\textsubscript{2} flux components

The RECCAP2 database 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\textsubscript{2} flux. The four simulations all start in preindustrial times and extend through 2018, however, the GOBMs used different definitions of “preindustrial” with simulations starting between 1765 and 1870, and the corresponding assumed pre-industrial CO\textsubscript{2} mixing ratios varying between 278 ppm and 286 ppm. Simulations A and C were forced with historically increasing CO\textsubscript{2}, whereas simulations B and D were forced with constant pre-industrial CO\textsubscript{2}. 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). Afterwards, simulations A and D were forced with these historical atmospheric reanalysis fields, whereas simulations B and C continued with the same constant atmospheric
reanalysis fields that were applied before 1948 or 1958. Thus, each simulation represents a different combination of the CO₂ flux components:

- Simulation A is forced with historical atmospheric reanalysis data and historically increasing CO₂, yielding:
  \[ F_{\text{SimA}} \approx F_{\text{ant}}^{\text{ss}} + F_{\text{ant}}^{\text{ns}} + F_{\text{nat}}^{\text{ss}} + F_{\text{nat}}^{\text{ns}} + F_{\text{riv-bur}}. \]  

- Simulation B is forced with the same repeated annual atmospheric forcing and constant pre-industrial CO₂ levels, yielding:
  \[ F_{\text{SimB}} \approx F_{\text{nat}}^{\text{ss}} + F_{\text{nat}}^{\text{riv-bur}}. \]  

- Simulation C is forced with a constant atmospheric forcing and historically increasing CO₂, yielding:
  \[ F_{\text{SimC}} \approx F_{\text{ant}}^{\text{ss}} + F_{\text{nat}}^{\text{ss}} + F_{\text{nat}}^{\text{riv-bur}}. \]  

- Simulation D is forced with historical atmospheric reanalysis data and constant pre-industrial CO₂ levels, yielding:
  \[ F_{\text{SimD}} \approx F_{\text{nat}}^{\text{ss}} + F_{\text{nat}}^{\text{ns}} + F_{\text{nat}}^{\text{riv-bur}}. \]

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 CO₂ (A, C) represent both natural and anthropogenic CO₂ fluxes, while simulations with pre-industrial CO₂ (B, D) represent only natural CO₂ fluxes.

The ocean physical and biogeochemical fields of the GOBMs were initialized with gridded observation-based estimates of ocean physics and biogeochemistry averaged over the last decades. The observation-based ocean DIC concentrations were thereby adjusted to represent pre-industrial DIC by removing the historical anthropogenic carbon uptake.

Four of the ten GOBMs considered here (FESOM-REcoM-LR, MOM6-Princeton, ORCA1-LIM3-PISCES, PlankTOM12) run the four simulations straight from these initial conditions without a pre-industrial spinup, while the remaining six (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 1586 years (overview table in DeVries et al. (in review)) using a repeated year of climatological atmospheric forcing and each model’s assumed pre-industrial atmospheric CO₂ with the goal to reach a near steady-state between the atmosphere and the ocean. Steady-state in this context refers to the state of a model under constant forcing, in which all multi-annual mean fluxes are time-invariant at the local scale and globally integrated zero. Few of the 6 models with spinup reach this objective, largely because of the spinup 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 bias in the simulated air-sea CO₂ fluxes (Griffies et al., 2016; Orr et al., 2017; Séférian et al., 2016). The models analyzed here have global CO₂ flux biases ranging between -0.35 and 0.17 PgC yr⁻¹, with a relatively small drift over time (Hauck et al., 2020). However, regionally, this effect can be more important. We call this bias in the sea-air CO₂ flux due to insufficient spinup and its drift
over time $F_{\text{drift+bias}}$. This $F_{\text{drift+bias}}$ 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 flux in one RECCAP2 simulation from the sea-air CO$_2$ flux in another RECCAP2 simulation, assuming that $F_{\text{nat\_riv-bur}}$ and $F_{\text{drift+bias}}$ are not affected by increasing atmospheric CO$_2$ or changing atmospheric forcing across the varying simulations and that the different flux components add up to the total flux without substantial non-linearities:

$$F_{\text{nat\_ss}} \approx F_{\text{SimC}} - F_{\text{SimB}}$$  (6)

$$F_{\text{nat\_ns}} \approx F_{\text{SimA}} - F_{\text{SimC}} + F_{\text{SimB}} - F_{\text{SimD}}$$  (7)

$$F_{\text{nat\_ns}} \approx F_{\text{SimD}} - F_{\text{SimB}}$$  (8)

The total air-sea CO$_2$ flux ($F_{\text{total}}$) can hence be estimated as follows:

$$F_{\text{total}} \approx F_{\text{SimA}} - F_{\text{SimB}} + F_{\text{nat\_ss}} + F_{\text{nat\_riv-bur}}$$  (9)

Globally, $F_{\text{nat\_ss}}$ is by definition zero, so that only $F_{\text{nat\_riv-bur}}$ has to be known for a GOBM-based estimate of $F_{\text{total}}$. Unfortunately, $F_{\text{nat\_riv-bur}}$ cannot be quantified from the here-used GOBM simulations because their set-ups were not designed to represent riverine input and/or sediment burial (see Section 3.1.1.). For the estimation of $F_{\text{total}}$ from GOBMs in RECCAP2, the observation-based estimate from (Regnier et al., 2022) was used in equation (9) as an approximation of global $F_{\text{nat\_riv-bur}}$ (i.e., 0.65±0.30 Pg C yr$^{-1}$), henceforth called $F_{\text{obs\_riv-bur}}$. 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.

Regionally, estimating $F_{\text{total}}$ from these simulations is more difficult, because the regional $F_{\text{nat\_ss}}$ is not zero as the ocean takes up and releases natural carbon regionally. Therefore, $F_{\text{total}}$ cannot be estimated as the difference between simulations A and B as this difference does not only remove $F_{\text{nat\_drift+bias}}$ and $F_{\text{nat\_riv-bur}}$, but also $F_{\text{nat\_ss}}$. Hence, we estimate regional $F_{\text{total}}$ from simulation A and accept the simulated regional $F_{\text{nat\_drift+bias}}$ and $F_{\text{nat\_riv-bur}}$ as inherent uncertainties. To still estimate $F_{\text{total}}$, we added an observation-based estimate of the regional $F_{\text{nat\_riv-bur}}$ ($F_{\text{obs\_riv-bur}}$) to the sea-air CO$_2$ fluxes from simulation A. This regional observation-based estimate of $F_{\text{obs\_riv-bur}}$ is derived from the estimated regional pattern of $F_{\text{nat\_riv-bur}}$ (Lacroix et al., 2020), which is scaled with a constant factor for all grid cells such that the global integral matches the postulated global value of $F_{\text{obs\_riv-bur}}$ of 0.65 Pg C yr$^{-1}$. Overall, the impossibility to disentangle the regional values of $F_{\text{nat\_ss}}$, $F_{\text{nat\_drift+bias}}$, and $F_{\text{nat\_riv-bur}}$ in the models and the uncertainties of the regional observation-based $F_{\text{obs\_riv-bur}}$ hence add additional uncertainty to the regional estimates of $F_{\text{total}}$. 
2.3 Observation-based estimates, their uncertainties and their usage for comparison with GOBMs

To compare the total sea-air CO$_2$ fluxes from the GOBMs with observation-based estimates, we utilize the RECCAP2 dataset of $p$CO$_2$ products, including AOML_EXTRAT, CMEMS-LSCE-FNN, CSIR-ML6, JenaMLS, JMA-MLR, MPI-SOMFFN, OceanSODA-ETHZ, UOEX_Wat20, and NIES-MLR3 (see table in DeVries et al. (in review)) for references and further details). These $p$CO$_2$ 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 only, i.e., for the period when all products provide estimates.

The simulated regional $F_{\text{nat}}^{ss}$ were compared to ocean inversion-based estimates (Mikaloff Fletcher et al., 2007). These rely on observations of interior ocean DIC, alkalinity, and nutrients to create a conservative DIC tracer where the anthropogenic concentration in each grid cell is subtracted following Gruber et al. (1996) as well as changes in the ocean interior DIC due to biological processes. In a second step, 10 ocean circulation models were used to determine the circulation pattern by injecting a dye tracer at the ocean surface at a constant rate. Finally, the circulation pattern which results in the best fit with the observations of the adjusted DIC tracer is utilized to determine $F_{\text{nat}}^{ss}$.

To constrain the simulated $F_{\text{ant}}^{ss}$, we used the mapped anthropogenic carbon storage between the years 1800 and 2002 from the GLODAPv2.2016b-product (Lauvset et al., 2016). This data-product is based on the TTD-method (Matear et al., 2003; Waugh et al., 2006) and henceforth referred to as TTD-estimate. It includes estimates of a mapping error, but a comprehensive error estimate containing observational, methodological, and mapping errors is not provided with the dataset. In lack of such an estimate, we utilized the error-estimate of ±29% for the $C_{\text{ant}}$-storage of the North Atlantic (Steinfeldt et al., 2009), which is a simplified and rather conservative error estimate (Khatiwala et al., 2013; Terhaar, Tanhua, et al., 2020). Additionally, the mapped $C_{\text{ant}}^{ns+ss}$-storage from the year 1800 to the year 1994 as well as that between the years 1994 and 2007 were quantified by Sabine et al. (2004) with the ocean tracer–based $\Delta C^*$ method (henceforth referred to as $\Delta C^*$-estimate) and by Gruber, Clement, et al. (2019) with the eMLR($C^*$)-method (henceforth referred to as eMLR($C^*$)-estimate), respectively. Uncertainties of the globally integrated estimates of both $C_{\text{ant}}^{ns+ss}$-storage estimates were provided when comparing these estimates to simulated values. We compared (changes of) anthropogenic carbon inventories between GOBMs and mapped TTD-, $\Delta C^*$- and eMLR($C^*$)-estimates (Section 3.3.1), respectively. As the mapped TTD-, $\Delta C^*$- and eMLR($C^*$)-estimates do not cover all Ocean basins (e.g., the Arctic Ocean and the Marginal Seas are not covered by the mapped TTD-estimate), the GOBM estimate is only integrated over grid-points that the associated mapped observation-based estimate covers. When referring to a comparison between TTD-, $\Delta C^*$- and eMLR($C^*$)-estimates and GOBM-estimate of interior ocean $C_{\text{ant}}$-storage then this excludes depth under 3000 m as well as the Arctic Ocean and the marginal Seas.

For the AMOC (here defined as maximum of the Atlantic meridional overturning streamfunction at 26°N), data from the RAPID-Meridional Overturning Circulation and Heatflux Array-Western Boundary Time Series array at 26°N were used (Frajka-Williams et al., 2021) to calculate the mean AMOC strength from 2005 to 2018. The measurement uncertainty of the annual AMOC values 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
The interfrontal sea surface salinity is the average sea surface salinity in the region limited by the polar front and the subtropical front and approximately describes the region where the upwelled circumpolar deep water is transformed into mode and intermediate water. Mean estimates and uncertainties were derived as described in Terhaar, Frölicher, et al. (2021) using gridded monthly climatologies of sea surface salinity and of sea surface temperature from the World Ocean Atlas 2018 (Locarnini et al., 2018; Zweng et al., 2018).

The volume of ventilated waters is defined as the volume of water south of 30°S with densities above the mean interfrontal sea surface density and below the mean interfrontal sea surface density plus 0.8 kg m$^{-3}$. The value of 0.8 kg m$^{-3}$ corresponds to approximately 2-3 times the area-weighted standard deviation of the monthly sea surface densities in the inter-frontal zone across the ensemble of ESMs used by Terhaar, Frölicher, et al. (2021). This density thus covers most of the denser water masses in the area that are relatively fast ventilated and excludes the small areas of very dense surface waters that very slowly ventilated a large amount of the deep ocean.

For comparisons of surface DIC and alkalinity between observation-based estimates and GOBMs, the observation-based 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) were used. As the gridded estimates of these three products are based on observations of surface ocean $p$CO$_2$ and alkalinity in space and time, we henceforth call them $p$CO$_2$/alkalinity products. Furthermore, gridded GLODAPv2 estimates of the same variables were also used (Lauvset et al., 2016), where DIC is normalized to the atmospheric $p$CO$_2$ of 2002. For comparison, output from the $p$CO$_2$/alkalinity products and GOBMs were averaged over the years 1986 to 2018, the longest time period available with the year 2002 in its center.

Additionally we compared the simulated and observation-based Revelle factor (Revelle and Suess, 1957), carbonate ion (CO$_3^{2-}$) concentrations, and the chemical surface ocean uptake capacity. CO$_3^{2-}$ acts as a buffer for the ocean carbon uptake (Broecker et al., 1979), which declines with increasing CO$_2$ uptake (Sarmiento & Gruber, 2006). The Revelle factor describes the overall uptake capacity of the ocean:

$$\text{Revelle} = (\Delta\text{DIC} / \text{DIC}) / (\Delta[p\text{CO}_2] / [p\text{CO}_2]).$$  \hspace{1cm} (10)

We re-arranged this equation to quantify the amount of additional carbon that the surface ocean can take up for a given increase in $p$CO$_2$ ($\Delta\text{DIC} / \Delta[p\text{CO}_2]$) and defined this to be the chemical uptake capacity:

$$\Delta\text{DIC} / \Delta[p\text{CO}_2] = \text{DIC} / (\text{Revelle} \times [p\text{CO}_2]).$$  \hspace{1cm} (11)

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 $p$CO$_2$/alkalinity products, and all GOBMs using mocsy2.0 (Orr & Epitalon, 2015), respectively, 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).

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., 2015) and each year to be independent as the moorings of the observational array are exchanged every year.
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, Kwiatkowski, et al., 2020; Terhaar, Frölicher, et al., 2021; Terhaar, Torres, et al., 2021; Terhaar et al., 2022).

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. When globally comparing the simulated $F_{total}$ of the GOBMs to that of the $pCO_2$ products, $F_{obs_{riv-bur}}$ has to be added to the GOBM estimate (see Section 2.2) and the relatively large 1-sigma uncertainty of $F_{obs_{riv-bur}}$ ($\pm 0.15$ Pg C yr$^{-1}$) substantially increases the uncertainty of the GOBM-derived estimate. For the global $F_{total}$ estimates from GOBMs, we will therefore provide both a combined uncertainty (standard deviation of GOBM ensemble and of $F_{obs_{riv-bur}}$) and a pure standard deviation that does not include the uncertainty of $F_{obs_{riv-bur}}$ and hence is a measure of model-based differences only. Excluding the uncertainty of $F_{obs_{riv-bur}}$ allows comparing the ensemble spread of estimates of GOBMs to that of the $pCO_2$ products. Regionally, the uncertainty of $F_{total}$ is only provided as the standard deviation across the GOBM ensemble, because regional uncertainties of $F_{obs_{riv-bur}}$ 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). The RECCAP2 biome-mask is a slightly modified version of the oceanic biomes of Fay & 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 a late starting date

To quantify the difference in the simulated anthropogenic carbon uptake from 1980 to 2018 due to different starting dates (see Section 2.2.3), it would be ideal to re-run all simulations that started later than 1765 from 1765 onwards. However, spinning-up several GOBMs with another pre-industrial $pCO_2$ and re-running the historical simulations from 1765 to 2018 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 and one in 1850, 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 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 $F_{\text{total}}$ (-1.41±0.24 Pg C yr$^{-1}$; here excluding uncertainties of $F_{\text{obs}}^{\text{riv-bur}}$) that is statistically indistinguishable from that estimated by the $p\text{CO}_2$-products (-1.71±0.26 Pg C yr$^{-1}$) (Table 1, Figure 1). In addition, the overall increasing trend is similarly represented by the two classes of estimates. Still, the difference of the long-term means of 0.30±35 Pg C yr$^{-1}$ (18±20% of the mean $p\text{CO}_2$-product estimate) is substantial. Moreover, the difference of annual mean fluxes between GOBMs and $p\text{CO}_2$-products varies with time, exceeding 20% of the average value of the $p\text{CO}_2$-products from 1985 to 1990, in 2009 and 2010, and from 2016 to 2018. 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 yr$^{-1}$ representing ~17% of their average $CO_2$-flux. Even larger differences are found on the regional scale (Figure 1b-f).

### Table 1: Ensemble mean estimate of global and regional $CO_2$-fluxes (Pg C yr$^{-1}$) by GOBMs and $p\text{CO}_2$ products. The GOBM uncertainty excludes the uncertainty of $F_{\text{obs}}^{\text{riv-bur}}$.

<table>
<thead>
<tr>
<th></th>
<th>Global</th>
<th>Atlantic</th>
<th>Pacific</th>
<th>Indian</th>
<th>Arctic</th>
<th>Southern</th>
</tr>
</thead>
<tbody>
<tr>
<td>GOBMs</td>
<td>-1.41±0.24</td>
<td>-0.23±0.15</td>
<td>-0.34±0.12</td>
<td>-0.10±0.06</td>
<td>-0.06±0.03</td>
<td>-0.73±0.31</td>
</tr>
<tr>
<td>$p\text{CO}_2$ products</td>
<td>-1.71±0.26</td>
<td>-0.37±0.06</td>
<td>-0.39±0.14</td>
<td>-0.13±0.04</td>
<td>-0.08±0.05</td>
<td>-0.74±0.07</td>
</tr>
</tbody>
</table>

Regionally, the time-averaged $F_{\text{total}}$ from 1985 to 2018 based on GOBMs and $p\text{CO}_2$-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 $p\text{CO}_2$ products (Table 1, Figure 1b). The difference in the Atlantic Ocean starts to increase around the year 2000, the same time when the $F_{\text{total}}$ estimates in the Arctic Ocean also start to diverge (Figure 1e). Furthermore, the GOBMs and the $p\text{CO}_2$ products do not show the
same decadal variability of $F_{\text{total}}$ in the Southern Ocean (Figure 1f). The inter-model ensemble spread of simulated $F_{\text{total}}$ is 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 $F_{\text{total}}$ 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 spread of $F_{\text{total}}$ in GOBMs stems from $F_{\text{ant}}^{\text{ss}}$ (Figure 2a, Table S1). Regionally, the spread of $F_{\text{total}}$ is dominated by the spread of the sum of $F_{\text{nat}}^{\text{ss}}$, $F_{\text{nat}}^{\text{riv-bur}}$, and $F_{\text{drift+bias}}$ in all basins but the Arctic Ocean (Figure 2b-d, Table S1). The second largest contributions to the model spread are $F_{\text{ant}}^{\text{ss}}$ and $F_{\text{nat}}^{\text{ns}}$. In the Arctic Ocean, the spread of the sum of $F_{\text{nat}}^{\text{ss}}$, $F_{\text{nat}}^{\text{riv-bur}}$, and $F_{\text{drift+bias}}$ and the spread of $F_{\text{nat}}^{\text{ns}}$ are of similar size (Figure 2e, Table S1). The relatively large importance of $F_{\text{nat}}^{\text{ns}}$ in the Arctic Ocean is mostly caused by sea ice decline, which is well represented in GOBMs, while the model spread in $F_{\text{nat}}^{\text{ss}}$ is caused by the inter-model differences in simulated $p$CO$_2$ under the melting sea ice (Yasunaka et al., in review).

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 $p$CO$_2$ products estimates, and propose adjustments to the GOBM results. A special focus will lie on the Atlantic Ocean, where the long-term mean difference between GOBMs and $p$CO$_2$ 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 $p$CO$_2$ products is largest.
Figure 1. Time series of global and regional sea-air CO$_2$ fluxes from 1980 to 2018 based on 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 pCO$_2$ products estimates (blue) for the a) 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.
Figure 2. Time series of sea-air \( CO_2 \) flux components globally and regionally from 1980 to 2018 based on GOBMs. The total sea-air \( CO_2 \) flux \( (F_{\text{total}}) \) integrated over each basin adjusted for the riverine-burial induced sea-air \( CO_2 \) flux (green) and the individual flux components from the GOBMs \( (F_{\text{ant}}^{\text{ss}} \text{ in red, } F_{\text{ant}}^{\text{ns}} \text{ in orange, } F_{\text{nat}}^{\text{ns}} \text{ in purple, and the sum of } F_{\text{nat}}^{\text{ss}}, F_{\text{nat}}^{\text{riv-bur}} \text{ and } F_{\text{drift+bias}} \text{ in brown}) \) 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 \( F_{\text{total}} \) does not include the uncertainty of the riverine adjustment.
3.1 Sea-air CO\textsubscript{2} fluxes in the steady state control simulation

3.1.1 Carbon fluxes from rivers and into sediments

The input of riverine carbon $F_{\text{nat}}^{\text{riv}}$ and the sedimentation of carbon $F_{\text{nat}}^{\text{bur}}$ is treated in various ways across the ensemble of GOBMs and varies from 0.00 Pg C yr\textsuperscript{-1} to 0.61 PgC yr\textsuperscript{-1} and from 0.00 Pg C yr\textsuperscript{-1} to 0.74 Pg C yr\textsuperscript{-1}, respectively (Table 2). The difference between $F_{\text{nat}}^{\text{riv}}$ and $F_{\text{nat}}^{\text{bur}}$ varies between -0.14 Pg C yr\textsuperscript{-1} and -0.54 Pg C yr\textsuperscript{-1} and is 0.10±0.23 Pg C yr\textsuperscript{-1} when averaged over the 8 GOBMs that provide all four simulations.

Table 2. Global ocean carbon fluxes (Pg C yr\textsuperscript{-1}) averaged from 1980 to 2018. Positive fluxes indicate fluxes out of the ocean, except for the land-sea river carbon fluxes. $F_{\text{nat}}^{\text{riv-bur}}$ was estimated as the difference between the land-sea river carbon flux and the burial in sediments, except for NorESM-OC1.2. $F_{\text{drift+bias}}$ was derived as the difference between $F_{\text{SimB}}$ and $F_{\text{nat}}^{\text{riv-bur}}$. The GOBM-ensemble values exclude MOM6-Princeton (see Section 2.1).

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

To estimate $F_{\text{nat}}^{\text{riv-bur}}$, we use the first-order assumption that $F_{\text{nat}}^{\text{riv-bur}}= F_{\text{nat}}^{\text{riv}}-F_{\text{nat}}^{\text{bur}}$ for all GOBMs except NorESM-OC1.2 (Table 2). This assumption ignores the potential influence of alkalinity and nutrient fluxes from riverine and sedimentation (Gao et al., 2023; Terhaar, Orr, Ethé,
et al., 2019) as we only have this information for NorESM-OC1.2. In that model, the carbon burial flux that is larger than the 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, which reduces the DIC storage capacity of the ocean. Overall, the alkalinity and carbon burial fluxes in NorESM-OC1.2 influence the sea-air CO2 flux in similar magnitude but with opposite signs so that $F_{\text{nat}}^{\text{riv-bur}}$ is almost zero (Table 2). With the adjusted $F_{\text{nat}}^{\text{riv-bur}}$ for NorESM-OC1.2, the multi-model mean $F_{\text{nat}}^{\text{riv-bur}}$ is $-0.03\pm 0.15$ Pg C yr$^{-1}$. In comparison, the model spread associated with $F^{\text{total}}$ is $0.24$ Pg C yr$^{-1}$. Although $F_{\text{nat}}^{\text{riv-bur}}$ does not directly affect the global estimation of $F^{\text{total}}$, it may substantially affect the regional estimates of $F^{\text{total}}$ and $F_{\text{ss}}^{\text{nat}}$.

3.1.2 Bias and drift in the sea-air CO2 flux due to incomplete spin-up

Across the ensemble of GOBMs, the approximated global $F^{\text{drift+bias}}$, quantified as the difference of $F^{\text{SimB}}$ and $F_{\text{nat}}^{\text{riv-bur}}$ (equation 3, Table 2), varies from $-0.35$ to $0.17$ Pg C yr$^{-1}$, with an ensemble mean of $-0.06\pm 0.18$ Pg C yr$^{-1}$. The model spread around $F^{\text{drift+bias}}$ is of similar order as the model spread associated with the global $F^{\text{total}}$ ($0.24$ Pg C yr$^{-1}$). We assume that this is mostly a consequence of a too short spinup and hence of models not being in a steady state, since the drift component in the sea-air CO2 flux from 1980 to 2018 (calculated as the trend of the global air-sea CO2 flux in simulation B) is less than $\pm 0.002$ Pg C yr$^{-2}$ for all GOBMs (Hauck et al., 2020). Although our estimation of $F^{\text{drift+bias}}$ 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 sufficiently long spin-up in each model to reach steady state may thus narrow down inter-model differences of regional $F_{\text{nat}}^{\text{ss}}$ and $F^{\text{total}}$.

3.1.3 Steady state natural sea-air CO2 flux

The mean $F^{\text{SimB}}$ estimates of the GOBMs from 1980 to 2018 (Figure 2) are $-0.11\pm 0.14$ Pg C yr$^{-1}$ for the Atlantic Ocean, $0.21\pm 0.13$ Pg C yr$^{-1}$ for the Pacific Ocean, $-0.06\pm 0.06$ Pg C yr$^{-1}$ for the Indian Ocean, and $-0.06\pm 0.01$ Pg C yr$^{-1}$ for the Arctic Ocean. In the Southern Ocean, the $F^{\text{SimB}}$ estimate of $-0.04\pm 0.27$ Pg C yr$^{-1}$ of the GOBMs is twice as uncertain as in the other basins. The relatively large uncertainty in the Southern Ocean may partly be the result of large inter-model differences in the simulated $F_{\text{nat}}^{\text{ss}}$ fluxes, as dynamically complex regions like the Southern Ocean are difficult to simulate (Sallée et al., 2013). Inter-model differences in $F^{\text{drift+bias}}$ likely also play a role for the uncertain $F^{\text{SimB}}$ 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 $F^{\text{drift+bias}}$ than in other ocean basins with less upwelling. The Southern hemisphere and especially the Southern Ocean are also the locations where the $F^{\text{drift+bias}}$ tends to be largest in Earth System Models (Séférian et al., 2016).

When comparing $F^{\text{SimB}}$ in the Southern Ocean to $F^{\text{SimB}}$ globally, a significant relationship ($r^2 = 0.62$, $p = 0.01$) with a slope of 1.03 can be identified (Figure 3a). This relationship suggests that global inter-model differences related to the sum of $F^{\text{drift+bias}}$ ($-0.06\pm 0.18$ Pg C yr$^{-1}$) and $F_{\text{nat}}^{\text{riv-bur}}$ ($-0.03\pm 0.15$ Pg C yr$^{-1}$) are indeed primarily stemming from the Southern Ocean, especially as...
such a relationship occurs in no other ocean basin (Figure 3b and Figure S2). Based this assumption, we subtract the sum of global ocean $F_{\text{drift+bias}}$ and $F_{\text{riv-bur}}$ (-0.09±0.22 Pg C yr$^{-1}$) for the GOBM-ensemble without MOM6-Princeton) from the Southern Ocean $F_{\text{SimB}}$ for each GOBM separately. This adjustment leads to an estimate of Southern Ocean $F_{\text{nat ss}}$ of 0.05 ± 0.18 Pg C yr$^{-1}$, 0.10 Pg C yr$^{-1}$ larger and of opposite sign than the non-bias adjusted average $F_{\text{SimB}}$ across all GOBMs and with a 33% smaller spread. The major part (>80%) of this adjustment is due to $F_{\text{drift+bias}}$. In the other basins, the regional $F_{\text{SimB}}$ does not seem to be significantly impacted by the sum of $F_{\text{drift+bias}}$ and $F_{\text{riv-bur}}$ across the GOBM ensemble or these fluxes cancel each other out. In these basins, we assume $F_{\text{SimB}}$ to be approximately equal to $F_{\text{nat ss}}$.

Our estimates of $F_{\text{nat ss}}$ can be compared to inverse estimates of $F_{\text{nat ss}}$ (Mikaloff Fletcher et al., 2007) (see also Section 2.3). These inverse estimates of $F_{\text{nat ss}}$ show larger uptake in the Atlantic (-0.24±0.08 Pg C yr$^{-1}$) and Pacific Ocean (-0.07±0.14 Pg C yr$^{-1}$), more outgassing in the Southern Ocean (0.44±0.11 Pg C yr$^{-1}$), and similar uptake in the Arctic (-0.02±0.01 Pg C yr$^{-1}$) and Indian Ocean (-0.12±0.04 Pg C yr$^{-1}$). 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). When changing the northern boundary of the Southern Ocean to 44°S, the adjusted regional $F_{\text{nat ss}}$ of the GOBMs changes to 0.27±0.19 Pg C yr$^{-1}$, still 0.18 Pg C yr$^{-1}$ smaller than the mean inverse-based estimate but within its uncertainties. Without the adjustment

**Figure 3.** Relationship between global and regional sea-air CO$_2$ fluxes of simulation B for 9 GOBMs. The relationship between sea-air CO$_2$ fluxes (averaged for 1980-2018, negative: into the ocean) of the global ocean and a) Southern Ocean and b) Atlantic Ocean is shown. Represented is the natural sea-air CO$_2$ flux plus a potential sea-air CO$_2$ flux bias due to an interior ocean drift and a sea-air CO$_2$ flux related to carbon fluxes from rivers and into sediments (simulation B). The dashed line indicates a linear fit and the shading the projection uncertainty with a 68% uncertainty interval. The same relationship for the other ocean basins is shown in Figure S2.
for $F^{\text{drift bias}}$ and $F^{\text{riv-bar}}$, the difference between the simulated and inverse-based estimate of $F^{\text{nat ss}}$ in the Southern Ocean are larger.

### 3.2 Non-steady state natural sea-air CO$_2$ flux

Averaged between 1980 and 2018, the GOBMs simulate an outgassing global $F^{\text{nat ns}}$ of 0.05±0.05 Pg C yr$^{-1}$. Here, we 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 $F^{\text{nat ns}}$ increase from 1980 to 2018 at a rate of 0.07±0.02 Pg C yr$^{-1}$ decade$^{-1}$ (Figure 2, Figure 4a). The tropical Pacific and the Indian section of the Southern Ocean are the main contributors to the trend towards stronger $F^{\text{nat ns}}$ carbon outgassing (Figure 4a, Figure 2). The average trend towards stronger outgassing of $F^{\text{nat ns}}$ is to a small part compensated by a trend towards non-steady uptake of natural CO$_2$ in the Northern Pacific and the Arctic Ocean (Figure 4a; Figure 2e; Yasunaka et al. (in review)). Across the model ensemble, large inter-model differences in the mean $F^{\text{nat ns}}$ 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 $F^{\text{nat ns}}$ of 0.16±0.03 Pg C yr$^{-1}$ 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 $F^{\text{nat ns}}$ 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 $F^{\text{nat ns}}$ is the dominant contributor to the inter-annual and decadal variability of $F^{\text{total}}$ in GOBMs and is globally 6 times larger than the variability in the climate-driven variability in the anthropogenic sea-air CO$_2$ fluxes ($F^{\text{ant ns}}$) and regionally 2 to 6 times larger (Figure 2). The simulated temporal variability of $F^{\text{total}}$ in the Pacific Ocean is driven by $F^{\text{nat ns}}$ (Figure 2c) and resembles the variability of $F^{\text{total}}$ in the $\rho$CO$_2$ 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.
Figure 4. Non-steady state natural sea-air CO$_2$ fluxes for 8 GOBMs. Maps of the a) multi-model trend and b) the multi-model 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 c) the multi-model mean and d) the multi-model standard deviation of the inter-annual variability of the natural non-steady state sea-air CO$_2$ flux (linear trend is removed).

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 8 GOBM ensemble resembles that of the TTD- and $\Delta C^*$-estimate (Figure 5, Figure S3) and 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). While the TTD- and $\Delta C^*$-based estimates and the here analyzed 8 GOBMs agree that the largest accumulation of $C_{ant}$ per surface area is located in the North Atlantic and at the northern limit of the Southern Ocean around 45$^\circ$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 3000 m depth that accumulated since preindustrial times. The simulated multi-model mean GOBM interior ocean \( C_{\text{ant}} \) is 83±15 Pg C in 1994, 22% (23 Pg C) lower than the \( \Delta C^* \)-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 from 1765 and 1870 (see Section 3.4.1) and biases in GOBM dynamics and biogeochemistry (see Section 3.4.5). 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 \( \Delta C^* \)-methodology might lead to an overestimation of \( C_{\text{ant}} \) in the upper water column and a negative bias in deeper waters (Matsumoto & Gruber, 2005).

As for the \( C_{\text{ant}} \)-storage since 1800, the spatial pattern of the simulated interior ocean \( C_{\text{ant}} \)-storage changes from 1994 to 2007 of the GOBMs resembles that of the eMLR\((C^*)\)-estimate (Figure 5, Figure S4). Over this recent period, the GOBM global model mean \( C_{\text{ant}} \)-storage change of 25±3 Pg C (only integrating over cells where \( C_{\text{ant}} \) estimates from the eMLR\((C^*)\) method exist) is also smaller than the eMLR\((C^*)\)-estimate, but only by approximately 20% (6 Pg C). The underestimation of the contemporary \( C_{\text{ant}} \)-storage change by GOBMs is likely smaller than the underestimation of \( C_{\text{ant}} \)-storage changes since 1800 because the late starting date of several GOBMs (Section 3.3.2) has a smaller effect on contemporary \( C_{\text{ant}} \)-storage changes. Regionally, differences between the GOBM mean and the eMLR\((C^*)\)-estimate (Figure 5) are most prominent in the Atlantic (Perez et al., to be submitted) and Southern Ocean (Hauck et al., to be submitted). The eMLR\((C^*)\)-estimate indicates an anomalously high rate of \( C_{\text{ant}} \)-change in the South Atlantic for the period 1994 to 2007 and an anomalously low rate of \( C_{\text{ant}} \)-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 \( C_{\text{ant}} \) 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 et al., to be submitted). However, the eMLR\((C^*)\)-estimate might also overestimate the strength of these anomalies, due to structural biases in the reconstructed changes of \( C_{\text{ant}} \) (Clement & Gruber, 2018; Gruber, Clement, et al., 2019).

Overall, the comparison of simulated and observation-based \( C_{\text{ant}} \) confirms that the GOBMs underestimate the oceanic storage of anthropogenic carbon and hence \( F_{\text{ant}^*} \) 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 \( C_{\text{ant}} \) storage in 1994 since the beginning of the industrialization and the simulated change in \( C_{\text{ant}} \) storage from 1994 to 2007 across the model ensemble (Figure S5) suggesting a bias in the model mean state that persists over centuries. In the following sections, we will analyze the model set-ups, and simulated circulation and biogeochemistry to identify reasons for the underestimation of \( F_{\text{ant}^*} \) by the GOBM ensemble.
Figure 5: Column inventories of historic and contemporary anthropogenic carbon storage changes, integrated from surface to 3000m depth. Visualised are a) e) i) observation-based estimates and related model-estimates based on 8 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) b) c) d) show results for $C_{ant}^{ss+ss}$ from the $\Delta C^*$-estimate for the period 1800-1994 and GOBM estimates from start date of each simulation to 1994, d) e) f) g) 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) j) k) l) show results for $C_{ant}^{ss+ss}$ 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.

3.3.2 Influence of pre-industrial atmospheric CO$_2$ mixing ratio on anthropogenic carbon uptake

The difference in the simulated sea-air CO$_2$ flux from 1980 to 2018 between the simulations starting in 1765 and those starting in 1850 is simulated by the EMIC Bern3D-LPX to be 0.04-0.06 Pg C yr$^{-1}$, depending on the ocean mixing strength (see Section 2.6 for details of this set-up). Regionally, most differences occur in regions of strong upwelling, such as the Southern Ocean (Figure 6b). 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 and more than 50% of this difference (9.8-13.7 Pg C) occurs after 1850.
Figure 6: Difference in anthropogenic sea-air CO\textsubscript{2} fluxes due to different starting dates in Bern3D-LPX. Maps of a) the anthropogenic sea-air CO\textsubscript{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 in 1850. Time series of c) the anthropogenic sea-air CO\textsubscript{2} flux the from simulations starting in 1850 with weak (blue), medium (orange), and strong (green) ocean mixing, and time series d) of the difference in the anthropogenic sea-air CO\textsubscript{2} flux the between simulations starting in 1850 and 1765 for the same ocean mixing strengths.

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. 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 (40x41 horizontal cells and only 3 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 adjustment term for these water masses.

Thus, the adjustment simulated by Bern3D-LPX for the air-sea CO\textsubscript{2} flux from 1980 to 2018 of 0.04-0.06 Pg C yr\textsuperscript{-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\textsubscript{2} fluxes
The large-scale pattern of the steady-state anthropogenic sea-air CO$_2$ flux ($F_{\text{ant}}^{\text{ss}}$) 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 $F_{\text{ant}}^{\text{ss}}$ and $C_{\text{ant}}$ in both North Atlantic and Southern Ocean (Figure 5, Section 3.3.1), similar to previous GOBMs (Orr et al., 2001) and ESMs (Frölicher et al., 2015; Goris et al., 2018; Terhaar, Frölicher, et al., 2021).

**Figure 7: Simulated mean and intermodel 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 8 GOBMs.

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 $C_{\text{ant}}$. It transports surface waters with high $C_{\text{ant}}$ (Pérez et al., 2013) and subsurface waters with low $C_{\text{ant}}$ (Ridge & McKinley, 2020) northwards. The subsurface waters outcrop in the subpolar gyre and are hence a sink of $C_{\text{ant}}$ (Ridge & McKinley, 2020). Both water masses are eventually transformed into deep water and transported southward. The AMOC is also the main driver of $F_{\text{ant}}^{\text{ss}}$ differences in the Atlantic across ensembles of ESMs from CMIP5 and CMIP6 (Goris et al., 2023; Terhaar et al., 2022), linking $F_{\text{ant}}^{\text{ss}}$ and the amount of $C_{\text{ant}}$ that was transported below 1000 m across these model ensembles (Goris et al., 2018, 2023).

Correlations between $F_{\text{ant}}^{\text{ss}}$ and (i) the AMOC at 26.5°N or (ii) the storage of $C_{\text{ant}}$ between 1000 m and 3000 m in the high latitude North Atlantic also occur across this ensemble of GOBMs and can be used to identify emergent constraints (Figure 8a,b). In combination with the respective observation-based estimates, the average annual Atlantic $F_{\text{ant}}^{\text{ss}}$ from 1980 to 2018 can be constrained from -0.39 ± 0.05 Pg C yr$^{-1}$ to -0.43 ± 0.06 Pg C yr$^{-1}$ when using the $C_{\text{ant}}^{\text{ss}}$ storage and to -0.42 ± 0.05 Pg C yr$^{-1}$ when using the AMOC. The constraints identify a common bias in the GOBMs towards too small AMOC strengths (mean underestimation of 18%) and $C_{\text{ant}}^{\text{ss}}$ storage below 1000m (mean underestimation of 22%), and hence Atlantic $F_{\text{ant}}^{\text{ss}}$ (mean underestimation of
8-10%, depending on the used constraint). Nevertheless, the uncertainties around the Atlantic $F_{\text{ant}}^\text{ss}$ estimate cannot be reduced due to the relatively large uncertainty of the observation-based estimate in case of the $C_{\text{ant}}^\text{ss}$ storage as well as the relatively weak but significant correlation between the AMOC and the Atlantic $F_{\text{ant}}^\text{ss}$ ($r^2 = 0.54$, $p = 0.04$). 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 $F_{\text{ant}}^\text{ss}$ than expected from the fit, respectively (Figure 8b). The correlation of the $C_{\text{ant}}^\text{ss}$ storage and $F_{\text{ant}}^\text{ss}$ is stronger ($r^2 = 0.84$, $p = 0.001$) because both variables are more directly related to each other and coherently affected by the late starting date. The relationships between Atlantic $F_{\text{ant}}^\text{ss}$ and (i) AMOC and (ii) $C_{\text{ant}}^\text{ss}$ storage between 1000 m and 3000 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 $C_{\text{ant}}^\text{ss}$ storage).

In the Southern Ocean, the magnitude of $F_{\text{ant}}^\text{ss}$ 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, i.e., 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 $F_{\text{ant}}^\text{ss}$ uptake (Terhaar, Frölicher, et al., 2021). As the density in the region of interest is almost entirely driven by the salinity (Supplement of Terhaar, Frölicher, et al. (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 $F_{\text{ant}}^\text{ss}$ and (i) the inter-frontal sea surface salinity, i.e., the mean surface salinity in the subtropical-polar frontal zone ($r^2=0.57$, $p=0.03$), and (ii) the volume of ventilated waters ($r^2=0.63$, $p=0.03$) (Figure 8c,d). 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, et al., 2021). Moreover, the biomes are partly defined based on the location of these fronts, so that 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 et al., to be submitted). The constraint with the sea surface salinity as predictor reduces the magnitude of $F_{\text{ant}}^\text{ss}$ in the Southern Ocean slightly from -0.74±0.09 Pg C yr$^{-1}$ to -0.72±0.08 Pg C yr$^{-1}$ (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, et al., 2021) between the sea surface salinity and $F_{\text{ant}}^\text{ss}$ can partly be explained by different starting dates as GOBMs with a late or early starting date have a smaller or larger absolute $F_{\text{ant}}^\text{ss}$ than expected from the linear fit between the mean surface salinity in the subtropical-polar frontal zone, respectively (Figure 8c).

A common starting date of 1765 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 $F_{\text{ant}}^\text{ss}$ because the scarcity of subsurface observations would have resulted in large uncertainties of the observational constraint.
While the here considered emergent constraints change the average annual $F_{\text{ant}}^{\text{ss}}$ from 1980 to 2018 in Atlantic and Southern Ocean only slightly, the influence of circulation biases on $F_{\text{ant}}^{\text{ss}}$ increases in magnitude with increasing atmospheric $F_{\text{ant}}^{\text{ss}}$. Therefore, the difference between constrained and unconstrained $F_{\text{ant}}^{\text{ss}}$ increases over time (Figure S6) and a GOBM ensemble with circulation biases will have smaller trends in $F_{\text{ant}}^{\text{ss}}$ and deviate from the true $F_{\text{ant}}^{\text{ss}}$ with time.

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) the Atlantic and c) the Southern Ocean, plotted against a) the Atlantic steady-state anthropogenic carbon storage between 1000 m and 3000 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) across GOBMs (green dots). 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 2.4 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).

3.3.5.2 Surface ocean carbonate chemistry

The $pCO_2$/alkalinity products suggest that the largest chemical surface ocean uptake capacity (defined here as $\Delta$DIC / $\Delta[pCO_2]$, 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
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 $p\text{CO}_2$/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 $\text{CO}_3^{2-}$ concentrations (Figure 9f), a high difference in surface 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 ~2300-2425 mmol m$^{-3}$) than ESMs (range of 2225-2415 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).

For the GOBMs, their globally different chemical uptake capacities do not explain their global differences in $F_{\text{ant}}^{ss}$ (Figure 9e), although studies with ESMs found such a relationship (Terhaar et al., 2022). Possible reasons for no emerging relationship between $F_{\text{ant}}^{ss}$ and the chemical uptake capacity, $\text{CO}_3^{2-}$, or the alkalinity across the GOBM ensemble are differences in $F_{\text{ant}}^{ss}$ due to different starting dates of the simulations (Section 3.3.2) and ongoing $F_{\text{drift+bias}}$. If a GOBM has a large negative or positive $F_{\text{drift+bias}}$, its upwelling waters have too low or high DIC, too high or low $\text{CO}_3^{2-}$, and hence a chemical uptake capacity that is too high or low, respectively. With time, the additional surface ocean DIC from $F_{\text{drift+bias}}$ reduces the chemical uptake capacity so that it is effectively smaller than the one expected from the theoretical chemical uptake capacity. Thus, $F_{\text{drift+bias}}$ adds considerable noise so that a potential relationship between the chemical uptake capacity and $F_{\text{ant}}^{ss}$ may not be identifiable. When considering only the four GOBMs with a longer spin-up than 1000 years, a relationship indeed emerges (Figure 9e-h).
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 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) 8 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 to 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$_2$ 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 for the variables on the respective x-axis.

3.4 Non-steady state anthropogenic sea-air CO$_2$ flux

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

Regionally, $F_{\text{ant}}^{\text{ns}}$ is substantially smaller than regional $F_{\text{nat}}^{\text{ns}}$ underlining the relatively minor importance of anthropogenic non-steady state fluxes compared to natural steady state fluxes. In the Southern Ocean, a strong negative trend in $F_{\text{ant}}^{\text{ns}}$ co-occurs in regions with strong positive trends in $F_{\text{nat}}^{\text{ns}}$ (Figure 4a). This suggests that both signals are related to stronger upwelling of circumpolar deep waters in most of the Southern Ocean with recent trends in climate as also discussed by Lovenduski et al. (2008) and Hauck et al. (to be submitted). This increased upwelling brings more old waters containing higher concentrations of $C_{\text{nat}}$ to the surface, enhancing the outgassing of $C_{\text{nat}}$. At the same time this exposes more waters to the surface with low concentrations of $C_{\text{ant}}$, causing an increase in $F_{\text{ant}}^{\text{ns}}$. In the North Atlantic subpolar gyre, the strong positive $F_{\text{ant}}^{\text{ns}}$ has a large model uncertainty associated with it, with some GOBMs showing a negative trend in $F_{\text{ant}}^{\text{ns}}$, 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 AMOC decline with climate-change 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 8 GOBMs. Maps of the a) the multi-model mean and b) the multi-model standard deviation of the linear trend in anthropogenic non-steady state sea-air CO$_2$ flux without the inter-annual variability (calculated by fitting a linear trend) averaged from 1980 to 2018, as well as maps of c) the multi-model mean and d) the multi-model standard deviation of the inter-annual variability (linear trend is removed).

4 Discussion and resulting recommendations

4.1 Spin-up and associated biases in the sea-air CO$_2$ flux

As not all GOBMs have been fully spun-up, globally integrated $F_{\text{drift+bias}}$ varies from -0.35 Pg C yr$^{-1}$ to 0.17 Pg C yr$^{-1}$ across the GOBM ensemble (-0.06±0.18 Pg C yr$^{-1}$ on average). $F_{\text{drift+bias}}$ does not directly affect our estimate of the global $F_{\text{total}}$ (based on equation (9)) as $F_{\text{drift+bias}}$ is removed when subtracting $F_{\text{SimB}}$ from $F_{\text{SimA}}$. In addition, other effects from a GOBM not being in steady-state owing to an insufficient spinup, such as biases in temperature, salinity, DIC, or alkalinity, and consequent biases in the circulation or chemical uptake capacity may still affect $F_{\text{total}}$. Regionally, $F_{\text{drift+bias}}$ directly affects $F_{\text{total}}$ because subtracting $F_{\text{SimB}}$ from $F_{\text{SimA}}$ removes not only $F_{\text{drift+bias}}$ but also $F_{\text{nat}}$, which is regionally not zero. To regionally estimate $F_{\text{total}}$ from a GOBM, one could hence rely either on $F_{\text{SimA}} - F_{\text{SimB}}$ and add an independent estimate of $F_{\text{nat}}$ (e.g., the inverse model estimate from Mikaloff Fletcher et al. (2007)), which comes with its own uncertainties, or rely on $F_{\text{SimA}}$ and treat the regional $F_{\text{drift+bias}}$ as an inherent uncertainty (as done here). Most of $F_{\text{drift+bias}}$ is likely located in the Southern Ocean and hence mostly affects the regional
estimate of $F_{\text{total}}$ in that ocean region. When assuming that $F_{\text{drift+bias}}$ is almost entirely located in the Southern Ocean (Section 3.3.1), $F_{\text{drift+bias}}$ could offset the total flux there (−0.7−0.8 Pg C yr$^{-1}$) by up to 50% in individual models and would increase the multi-model mean by ~13%.

A model-by-model analysis would be necessary to determine the extent of the spinup related bias and drift in each GOBM and the necessary length of the spinup for a GOBM to reach steady state. Depending on the difference between the model’s steady-state and the initialization, the necessary length of the spinup may vary between individual GOBMs (Gürses et al., 2023) (see also Figure 11). Such a model-by-model assessment of the necessary spinup length would include the assessment of different variables in different regions and depth-ranges and exceeds the scope of this study. A comparison between the number of simulated years before the start of the analysis period of each GOBM and the $F_{\text{drift+bias}}$ (Figure 11) suggests that a short spin-up is often insufficient to reduce $F_{\text{drift+bias}}$ (Griffies et al., 2016; Orr et al., 2017; 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 $F_{\text{total}}$ 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 model’s physics, biology, and carbonate chemistry.

Figure 11. Estimated global sea-air CO$_2$ bias fluxes related to the models not being in steady-state for 9 GOBMs 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-flux ($F_{\text{bias}}$) is determined as specified in Section 3.2, Table 2. The spin-up ORCA025-GEOMAR was branched from a previous spin-up from the same model but with a coarser resolution.

4.2 Riverine and sediment fluxes

The GOBMs differ strongly in their representation of the riverine and sediment fluxes of carbon, nutrients and alkalinity, ranging from models without such fluxes to models with that attempt to resolve these fluxes explicitly. According to our approximation, none of the GOBMs simulates a resulting riverine and sediment flux-driven $F_{\text{nat riv-bur}}$ comparable to the observation-based $F_{\text{obs riv-bur}}$ of -0.65±0.30 Pg C yr$^{-1}$ (Regnier et al., 2022). The different representation of the
riverine and sediment fluxes in the GOBMs thus represent an important inherent uncertainty of the model-simulated regional sea-air CO$_2$ fluxes and the global natural sea-air CO$_2$ fluxes. Global GOBM-estimates of $F_{\text{total}}$ are however unaffected by $F_{\text{nat}}^{\text{riv-bur}}$ if equation (9) is used as the simulated $F_{\text{nat}}^{\text{riv-bur}}$ is removed when subtracting $F_{\text{SimB}}^{\text{obs}}$ from $F_{\text{SimA}}^{\text{obs}}$ and replaced by the observation-based estimate of riverine and sediment fluxes. Apart from riverine carbon and alkalinity fluxes, an inadequate representation of riverine nutrient fluxes can also affect all components of the sea-air CO$_2$ fluxes 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). However, estimates of the impact of changing riverine carbon, alkalinity and nutrient fluxes 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). More research and model development is urgently needed to better represent the riverine and sediment fluxes in GOBMs to allow for a less uncertain quantification of global and regional sea-air CO$_2$ fluxes. An accurate observation-based estimate of the global riverine and burial derived sea-air CO$_2$ flux is necessary to estimate global $F_{\text{total}}$ for GOBMs without these fluxes. Despite large efforts over the last years (Lacroix et al., 2021; Regnier et al., 2022; Resplandy et al., 2018), the most recent observation-based estimate of the global $F_{\text{obs}}^{\text{riv-bur}}$ of $-0.65 \pm 0.30$ Pg C yr$^{-1}$ (Regnier et al., 2022) still has large uncertainties (~45%) that even exceed the simulated inter-model standard deviation of $F_{\text{total}}$ before accounting for $F_{\text{obs}}^{\text{riv-bur}}$ ($\pm 0.24$ Pg C yr$^{-1}$).

Regionally, the uncertainties of $F_{\text{obs}}^{\text{riv-bur}}$ are even larger than globally. Across RECCAP2 chapters, the local distribution of $F_{\text{obs}}^{\text{riv-bur}}$ is derived from Lacroix et al. (2021) (see Section 2.3.3), suggesting a strong riverine-burial-induced carbon outgassing in the Atlantic Ocean (0.27 Pg C yr$^{-1}$) and a relatively weak riverine-burial-induced carbon outgassing in the Southern Ocean (0.04 Pg C yr$^{-1}$). Contrarily, an older estimate by Aumont et al. (2001) suggests a smaller $F_{\text{obs}}^{\text{riv-bur}}$ in the Atlantic Ocean and a larger $F_{\text{obs}}^{\text{riv-bur}}$ in the Southern Ocean. One reason for this difference might be that Lacroix et al. (2021) quantify $F_{\text{obs}}^{\text{riv-bur}}$ as the difference between a simulation with observation-based riverine fluxes of carbon 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 removal of the artificial surface ocean carbon and nutrient input may override the riverine signal, especially in regions far away from river deltas such as the Southern Ocean. The artificial carbon input in the reference simulation would also explain why the global estimate of $F_{\text{obs}}^{\text{riv-bur}}$ of Lacroix et al. (2021) is smaller than other existing estimates (Aumont et al., 2001; Regnier et al., 2022; Resplandy et al., 2018). Another reason for the difference of the spatial distribution of $F_{\text{obs}}^{\text{riv-bur}}$ between Lacroix et al. (2021) and Aumont et al. (2001) is the assumption of the lability of riverine organic matter, which is lower in Lacroix et al. (2021). 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. (2021) would instead occur in the Southern Ocean, $F_{\text{total}}$ in the Atlantic Ocean would double. Hence, more refined estimates of the lability of organic matter and its effect on $F_{\text{obs}}^{\text{riv-bur}}$ are crucial to better constrain the total sea-air CO$_2$ flux and regional anthropogenic carbon sink estimates.

4.3. Starting date and pre-industrial CO$_2$
The different pre-industrial atmospheric CO$_2$ in each GOBM introduces a difference in the simulated anthropogenic carbon flux between the GOBMs (Section 3.3.2). We compared two estimates for the impact of a later starting date on the anthropogenic carbon fluxes, which suggest that a later starting date leads to a global underestimation of 0.04-0.06 Pg C yr$^{-1}$ (3-5% of $F_{total}$) for the period 1980-2018. However, this underestimation of 0.04-0.06 Pg C yr$^{-1}$ is highly uncertain and possibly underestimated by about 40%.

To avoid the need of an estimate of the underestimation and the uncertainties that come with it, our recommendation would be to start all simulations in 1765 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), and Mikaloff Fletcher et al. (2006)). While this necessitates to perform up to 85 more years per simulation, the cost of running GOBMs in hindcast mode is much smaller than the cost of fully-coupled Earth System Models and computational constraints should thus not represent a major bottleneck.

4.4 Circulation biases

Previously identified relationships in ESMs between the AMOC and the North Atlantic $F_{ant}^{ss}$ (Goris et al., 2018) and the inter-frontal sea surface salinity and the Southern Ocean $F_{ant}^{ss}$ (Terhaar, Frölicher, et al., 2021) could also be identified in this GOBM ensemble. Overall, the considered GOBMs underestimate the strength of the AMOC (on average by 3.1±5.2 Sv at 26.5°N) and hence $F_{ant}^{ss}$ in the Atlantic and slightly overestimate the inter-frontal sea surface salinity (on average by 0.05±0.12) and hence $F_{ant}^{ss}$ in the Southern Ocean, though the resulting improvements of both constrained $F_{ant}^{ss}$ estimates are small for the ensemble average.

The on average relatively good agreement of the simulated and observed sea surface salinity between the polar and subtropical fronts in the Southern Ocean is a direct consequence of the forcing with atmospheric observation-based temperatures from reanalysis products such that the location of the fronts is well presented by the models. In addition, some of the GOBMs also restore the salinity at the ocean surface towards observed salinities. Despite this, some GOBMs still overestimate the salinity substantially.

The AMOC strength at 26°N, however, differs significantly across our considered GOBMs with its multi-model mean being negatively biased. 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, only CESM-ETHZ has an extraordinarily small AMOC, which was improved in a later simulation set-up version. This led to larger $F_{ant}^{ss}$ uptake in the Atlantic (see Figure 8b). The substantial change in the AMOC from 3.5 to 14.8 Sv in CESM-ETHZ due to a different sea surface salinity restoring timescale, i.e., a different artificial salinity flux across the air-sea interface, highlights the strong sensitivity of the ocean circulation to atmospheric fluxes.

The here used emergent constraints provide relatively robust relationships between circulation features and carbon fluxes, which were tested across the CMIP5 and CMIP6 ensembles and the here used GOBM ensemble. 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). While the best estimates of $F_{\text{ant}ss}$ in the Atlantic and Southern Oceans have changed the original estimate by less than 10% here, other model ensembles might have larger biases and changes in $F_{\text{ant}ss}$ might hence be different. The relatively small reduced uncertainty in both regions (<11%) is likely due to weaker correlations due to different pre-industrial $p\text{CO}_2$ in the ensemble of GOBMs, which can relatively easily be improved following our recommendation of starting all simulations in 1765. In the long-term, we recommend improving the representations of key ocean circulation metrics in the GOBMs.

4.5 Ocean biogeochemistry

The globally averaged chemical uptake capacity does not show a strong relationship with globally integrated $F_{\text{ant}ss}$ across the GOBM ensemble (Figure 9e) although such a relationship was found across an ensemble of ESMs (Terhaar et al., 2022). Here, the relationship might be blurred by other processes that are influencing the simulated $F_{\text{ant}ss}$, namely circulation biases, different starting dates and bias due to different spin-up length. Accounting for the influence of the bias in circulation on $F_{\text{ant}ss}$ (Section 3.4.5.1), i.e., increasing $F_{\text{ant}ss}$ in the North Atlantic for GOBMs with a too small AMOC, supports the relationship, but does not lead to a tighter relationship. If only GOBMs with a spin-up above 1000 years were considered, a linear relationship between $F_{\text{ant}ss}$ and the chemical uptake capacity emerges (Figure 9e) that resembles the same relationship across ESMs (Terhaar et al., 2022). However, the small number of considered GOBMs and the range of observation-based estimates of the chemical uptake capacity does not allow to exploit such a potential relationship yet. Eventually, only a GOBM ensemble with all models being spun-up to steady-state and better constrained observation-based estimates would allow drawing such conclusions more robustly.

4.6 Gap between observation-based estimates and GOBMs

For the period 1985-2018, our analysis identifies a gap in $F_{\text{total}}$ of 0.30 Pg C yr$^{-1}$ between surface $p\text{CO}_2$ products (-1.71±0.26 Pg C yr$^{-1}$) and GOBMs (-1.41±0.28 Pg C yr$^{-1}$; uncertainty includes the ±0.15 Pg C yr$^{-1}$ 1-sigma uncertainty of the $F_{\text{obs}riv-bur}$ estimate). The GOBM underestimation of 0.30±0.38 Pg C yr$^{-1}$ (~18% of the $F_{\text{total}}$ of the $p\text{CO}_2$ products) can partially be explained by the late starting date of the GOBM simulations, circulation biases, and potential biogeochemical biases in the GOBMs. In addition, comparisons of the simulated $C_{\text{ant}}$ storage since the beginning of the industrialization and over recent years from 1994 to 2007 to observation-based estimates also suggests that the GOBMs underestimate $F_{\text{ant}ss}$ by 20-30%. Apart from our identified average gap between GOBM and $p\text{CO}_2$ product estimates of the ocean carbon sink, we confirm that the trends in the ocean sink since 2000 also differ globally and regionally (Friedlingstein et al., 2022; Hauck et al., 2020). Although these different trends suggest a divergence between GOBM estimates and $p\text{CO}_2$ products in recent years (Figure 1a), an increase in $F_{\text{total}}$ by around ~20% in each year accounting for an underestimation of the anthropogenic steady-state flux would change this perception. The difference in $F_{\text{total}}$ would not appear as a divergence of both estimates since 2000 but as a change from an underestimation of $F_{\text{total}}$ by the
$p\text{CO}_2$ products to an overestimation. Nevertheless, the growth rates of $F^\text{total}$ are different between GOBMs and $p\text{CO}_2$ products and uncertainties remain of how the ocean sink evolves.

Regionally, different trends in $F^\text{total}$ between GOBMs and $p\text{CO}_2$ products seem to be driven by a mismatch in the temporal evolution of the Southern Ocean carbon sink, and an increasing gap between both estimates in the Atlantic (Figure 1). In the Southern Ocean, the $p\text{CO}_2$ product estimate of the Southern Ocean carbon sink suggested that the variability before 2000 is mainly due to decadal variations (Gruber, Landschützer, et al., 2019; Keppler & Landschützer, 2019; Landschützer et al., 2015; McKinley et al., 2017). Since 2000, the estimate of the $p\text{CO}_2$ products of the Southern Ocean carbon flux has been moving toward more uptake. While this ongoing increase in uptake based on the $p\text{CO}_2$ 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 $p\text{CO}_2$-based and GOBM-based estimates for unknown reasons. Moreover, it remains an open question if differences between both estimates are due to the erroneous models or the extrapolation of sparse observations with temporal aliasing.

The increasing gap in the Atlantic after 2000, however, appears to result from a smaller $F^\text{total}$ trend in GOBMs than in $p\text{CO}_2$ products. This smaller-than-observed trend in GOBMs 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 show a biased-small AMOC, whose influence on $F^\text{ant,ss}$ increases with increasing atmospheric $CO_2$ (Section 3.3.3.1; Figure S6). Perez et al. (to be submitted) show that the disagreement in Atlantic $F^\text{total}$ trends between GOBMs and $p\text{CO}_2$ products is especially large in the subpolar North Atlantic. This relates well to our finding about AMOC-biases as the influence of AMOC-biases on $F^\text{ant,ss}$ is potentially highest in the subpolar gyre where subsurface waters low in $C^\text{ant}$ outcrop. Furthermore, a study with ESMs has shown that AMOC-biases are strongly correlated to SST-biases in the North Atlantic (Wang et al., 2014). While we did not analyze SST biases in the North Atlantic, Rodgers et al. (in review) found that the seasonal cycle of $p\text{CO}_2$ in the subpolar Atlantic is thermally driven in the GOBMs while that of the $p\text{CO}_2$-products is non-thermally driven. This might lead to the $F^\text{total}$ of the GOBMs being more sensitive to warming (Goris et al., 2018), which may contribute to the increasing gap between GOBMs and $p\text{CO}_2$-products with time. However, the magnitude of these contributions is unclear and remains to be identified.

4.7 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 $F^\text{total}$ results from $F^\text{nat,ns}$, i.e., the sea-air flux of natural carbon due to climate variability and climate change. Globally, $F^\text{nat,ss}$ is also an important flux component as it allows comparing the estimated ocean carbon sink from surface ocean $p\text{CO}_2$ products, which quantify $F^\text{ant,ss}$, $F^\text{ant,ns}$, $F^\text{nat,ns}$, and $F^\text{obs,riv-bur}$ (Friedlingstein et al., 2022) to observation-based estimates of the interior ocean change of $C^\text{ant}$ (Gruber, Clement, et al., 2019), which quantifies only changes in $F^\text{ant,ss}$, and $F^\text{ant,ns}$.
Previous approximations estimated the global $F_{\text{nat}}^{\text{ns}}$ from 1994 to 2007 to be $5\pm3$ Pg C (Gruber, Clement, et al. (2019); based on observation-based estimates of anthropogenic carbon fluxes storage changes and surface ocean fluxes), to be $1.3$Pg C (Friedlingstein et al. (2022), based on GOBMs) and to be $1.6\pm0.8$ Pg C (Terhaar et al. (2022), based on ESM simulations). The GOBMs here estimate a $F_{\text{ant}}^{\text{ns}}$ of $1.6\pm0.8$ 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 $p$CO$_2$ products in the Pacific Ocean, where most of the inter-annual variability is located, and differs in the Southern Ocean, where $p$CO$_2$ products suggest a strong decadal variability before 2000 and a different trend after 2000 (Gloege et al., 2021; Gruber, Landschützer, et al., 2019; Landschützer et al., 2015) (Figure 1). However, the sparse observations in the Southern Ocean pose a challenge for the observation-based estimates. For example, Gloege et al. (2021) showed that the 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 these differences in variability between between GOBMs and $p$CO$_2$ products in the Southern Ocean 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 $F_{\text{nat}}^{\text{ns}}$, the largest $F_{\text{ant}}^{\text{ns}}$ 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 $F_{\text{ant}}^{\text{ns}}$ in the Southern Ocean could be explained by additional free ocean surface due to climate change, which can thus take up more $C_{\text{ant}}$ or by more upwelling of old water with low $C_{\text{ant}}$ content (Le Quéré et al., 2007), which can also take up more $C_{\text{ant}}$. Both processes would lead to partial compensation by $F_{\text{nat}}^{\text{ns}}$ fluxes (Hauck et al., to be submitted; Lovenduski et al., 2008), with either more natural carbon being upwelled to the surface or more $C_{\text{nat}}$ being released with reduced ice cover.

4.8 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) and Hauck et al. (2020)). Hauck et al. (2020) found that GOBMs on average overestimate the observed $p$CO$_2$ 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 sink primarily because of circulation biases. The late-starting date and biases in the chemical uptake capacity in models also tend to enhance this underestimation. Fay & McKinley (2021) tested how well GOBMs resemble the $p$CO$_2$ 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 & McKinley (2021), such as the application of the local riverine adjustment by Lacroix et al. (2021), not accounting for each models’ simulated regional $F_{\text{nat riv-bur}}$ and that an area-weighted repartitioning $F^{\text{drift+bias}}$ over the entire ocean surface is valid. However, the local riverine adjustments come with large uncertainties (Section 4.2) and our analysis suggests that $F^{\text{drift+bias}}$ and $F_{\text{nat riv-bur}}$ are not evenly distributed. These adjustments affect the regional $F_{\text{total}}$ and don’t allow for robust simulated estimates of the regional $F_{\text{total}}$. Therefore, constraining the global $F_{\text{total}}$ with regional $F_{\text{total}}$ appears to be prone to large uncertainties and we recommend rather using underlying physical and biogeochemical processes for such constraints.

5 Conclusions

Our analysis of GOBMs helps to explain inter-model differences and differences between $pCO_2$-products and ocean biogeochemistry models estimates of the ocean carbon sink (DeVries et al., in review; Friedlingstein et al., 2022). These differences can be divided into (i) differences in the simulation set-ups, i.e., starting year and model spin-up, (ii) dynamical differences, i.e., model physics and biogeochemistry, and (iii) differences in boundary fluxes across the land-sea and sea-sediment interfaces.

The differences in the simulation set-ups can be resolved relatively easily by (a) using the CO$_2$ mixing ratio from 1765 as pre-industrial value and branching the historical simulation from the pre-industrial control simulation in 1765 and (b) increasing the spin-up period to reduce the uncertainty of the simulated $F_{\text{total}}$ in relation to model drift and allows to pinpoint weaknesses of the GOBMs and relationships across the GOBMs which are more apparent in steady-state.

Although one might suspect that an increasing spin-up would cause models to diverge from observations, we have found no evidence for this in this GOBM ensemble (Figure 8 and 9). Starting simulations in 1765 is an attractive option as 85 years of simulation may remove a global bias that is at least 0.04-0.06 Pg C yr$^{-1}$ in simulations that started in 1850 (underestimation of the sink). We here recommend using 1765 and not 1800 as in the TTD and $\Delta C^*$ estimates as the difference between atmospheric pCO2 in 1765 and 1800 already has a substantial effect on the ocean carbon sink until today (Bronselaer et al., 2017). The bias due to a too short spin-up is already accounted for on a global level through subtraction of the flux of the control simulation and hence does not affect estimates of the global carbon sink, such as the Global Carbon Budget estimate (Friedlingstein et al., 2022). However, a too short spin-up does impact regional flux estimates, particularly in the Southern Ocean. Moreover, where the models not being in steady-state also influences 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.

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 interfrontal sea surface salinity) provide robust relationships to adjust simulated anthropogenic carbon fluxes for these two key processes while these presentations are not improved yet. Our results show that the 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). Both 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 set-up 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 long-term, we recommend more complex adjustments within the set-ups of the GOBMs to reduce these biases.

The relatively poor representation of riverine and burial fluxes introduces another uncertainty to the simulated sea-air CO$_2$ fluxes. 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 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 total sea-air CO$_2$ fluxes. Improving the representation of these fluxes and their underlying processes is thus of importance to better understand the regional ocean carbon sink.

As 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), it remains challenging to compare the modeled estimates to the observation-based estimates of the ocean carbon sink. 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 $F_{\text{total}}$ or has to be subtracted from the $pCO_2$ products to be able to compare these estimate 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) and Lacroix et al. (2020)), the regional distribution and temporal variability of these fluxes still remains highly uncertain and renders a comparison between simulated and observation-based estimates of the ocean carbon sink complicated.

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 (late starting date, circulation biases, biogeochemical biases, biases in $C_{\text{ant}}$ storage) suggest that the GOBMs underestimate the ocean carbon sink on average. 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. Although explanations exist for the difference in the long-term mean carbon sink, the difference between the growth rates of the ocean carbon sink since 2000 globally, and in the Southern and Atlantic Oceans remains unresolved.

Overall, the model evaluation has helped to give recommendations for the set-up 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 in 1765, and increase the spin-up 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 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.

Acknowledgments

The authors have no conflicts of interest. J. Terhaar acknowledges funding from the Woods Hole Oceanographic Institution Postdoctoral Scholar Program, and the Swiss National Science Foundation under grant #200020_200511. 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. 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 number VH-NG-1301). 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. J. Terhaar thanks the science team and the crew of the R/V Sikuliaq for their great kindness, professional working atmosphere, great food, and moral support during the cruise in the Arctic Ocean in November, which provided an excellent environment for advancing this work.

Open Research

All of the RECCAP2 data will be made available in a public repository before publication.

References


https://doi.org/10.1038/ngeo1391

https://doi.org/https://doi.org/10.1029/2001GB001415

https://doi.org/https://doi.org/10.1029/2004GB002397


https://doi.org/https://doi.org/10.1029/2019MS002043


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

Jens Terhaar¹²³⁴, Nadine Goris⁴, Jens D. Müller⁵, Tim DeVries⁶⁷, Nicolas Gruber⁵, Judith Hauck⁸, Fiz F. Perez⁹, Roland Séférian¹⁰

¹Department of Marine Chemistry and Geochemistry, Woods Hole Oceanographic Institution, Woods Hole, 02543, Massachusetts, USA
²Climate and Environmental Physics, Physics Institute, University of Bern, Switzerland
³Oeschger Centre for Climate Change Research, University of Bern, Switzerland
⁴NORCE Climate & Environment, Bjerknes Centre for Climate Research, Bergen, Norway
⁵Environmental Physics, Institute of Biogeochemistry and Pollutant Dynamics, ETH Zurich, Zürich, Switzerland
⁶Department of Geography, University of California, Santa Barbara, Santa Barbara, CA, USA
⁷Earth Research Institute, University of California, Santa Barbara, Santa Barbara, CA, USA
⁸Alfred-Wegener-Institut, Helmholtz-Zentrum für Polar- und Meeresforschung, Am Handelshafen 12, 27570 Bremerhaven, Germany
⁹Instituto de Investigaciones Marinas (IIM), CSIC, Vigo, Spain. (0000-0003-4836-8974)
¹⁰CNRM (Université de Toulouse, Météo-France, CNRS), Toulouse, France

Corresponding author: Jens Terhaar (jens.terhaar@unibe.ch)
†Nadine Goris and Jens Terhaar should be considered joint first authors.

Key Points:

• The simulated CO₂-uptake by global ocean biogeochemistry models in RECCAP2 is systematically lower than observation-based estimates
• This underestimation is caused by the set-up of the RECCAP2-simulations as well as biases in surface chemistry and ocean circulation.
• Concrete steps forward are proposed to improve simulations of the ocean carbon sink by global ocean biogeochemistry models
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 $p$CO$_2$ 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 the (i) model set-up, such as the length of the spin-up, the starting date of the simulation, and carbon fluxes from rivers and into sediments, (ii) the ocean circulation, such as Atlantic Meridional Overturning Circulation and Southern Ocean mode and intermediate water formation, and (iii) the oceanic buffer capacity. Our analysis suggests that the 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 set-up prevents a robust assessment. For simulations of the ocean carbon sink, we recommend in the short-term to (1) start simulations in 1765, when atmospheric CO$_2$ started to increase, (2) conduct a sufficiently long spin-up such that the GOBMs reach steady-state, and (3) 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) with regards to their simulated CO$_2$ uptake from 1980 to 2018. We focus our analysis on the simulation set-up from the Global Carbon Budget initiative and the GOBMs that are used in the current version of the Global Carbon Budget. We find that the simulated CO$_2$ uptake by GOBMs is systematically lower than that of observation-based estimates and that the estimates differ substantially between GOBMs. We identify several reasons for this underestimate, relating to the set up of the simulations as well as the set up of the GOBMs themselves. For the set-up of the simulations, we find that not all GOBMs had the same starting year and the same initial stability, while the set up of the GOBMs themselves showed that the majority of GOBMs underestimate the large scale ocean circulation in the Atlantic and do not provide the necessary output for evaluation of their land-ocean interface. Based on our evaluation, we give recommendations for the set-up of follow up studies.

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; Terhaar et al., 2022), thereby reducing the growth of atmospheric CO$_2$ and slowing down global warming (IPCC, 2021). 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 CO₂ concentrations is known as anthropogenic carbon (Cₘₐₚ; Sarmiento et al., 1992), while the DIC that existed prior to the start of the industrial revolution is called natural carbon (Cₙₐt). 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 CO₂ fluxes (Joos et al., 1999; McNeil & Matear, 2013; Le Quéré et al., 2000), 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 Cₘₐₚ flux is controlled by the rate of Cₘₐₚ transport from the surface ocean to the deep ocean, which depends on the concentration of Cₘₐₚ 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 Cₙₐt is primarily controlled by the interaction of surface heating or cooling affecting the solubility of CO₂ in seawater and transport and mixing, and biological processes of photosynthesis, respiration, and CaCO₃ production (Sarmiento & Gruber, 2006). While there is agreement on these drivers for Cₘₐₚ and Cₙₐt 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 CO₂ (pCO₂ products) and hindcast simulations from global and regional ocean biogeochemistry models (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 pCO₂ products, the compiled database of RECCAP2 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 CO₂ 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 pCO₂ products (Chau et al., 2022; Gregor et al., 2019; Gregor & Gruber, 2021; 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 with the multi-model mean simulated net oceanic carbon sink reported by the Global Carbon Budget being consistently less negative (lower uptake) than the mean estimate of the pCO₂-products (1990s: -1.91±0.25 Pg C yr⁻¹ in models vs -2.14±0.34 Pg C yr⁻¹ for pCO₂ products, 2000s: -2.05±0.27 Pg C yr⁻¹ vs -2.34±0.21 Pg C yr⁻¹, and 2010s: -2.42±0.29 Pg C yr⁻¹ vs -3.02±0.22 Pg C yr⁻¹; Friedlingstein et al., 2022). The difference between the models and pCO₂ products in the 2010s is around half as large as the annual CO₂ 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; Terhaar et al., 2022) and Southern Ocean (Bourgeois et al., 2022; Fu et al., 2022; Terhaar, Frölicher, et al., 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 set-up 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 around 1765 (Bronsema et al., 2017; Terhaar, Orr, Gehlen, et al., 2019), leading to a too low ocean carbon uptake and storage. Associated with the set-up 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 $p$CO$_2$ 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 simulate their cycling through biogeochemical processes (primary production, grazing, remineralization, etc.) (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 8 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 also explain the reasons for its exclusion there, we exclude it from all GOBM results in terms of multi-model mean and standard deviation.

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 an overview table in DeVries et al. (in review). 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 simulations, the salinity restoring timescale was two 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 timescale in the Southern Ocean better captures sea-ice related fluxes that are not well represented in the atmospheric forcing fields. This change in the salinity restoring led to an improvement of the modeled overturning circulation, not only in the Southern Ocean, but also in the North Atlantic, where the previously very weak Atlantic Meridional Overturning Circulation (AMOC) increased from 3.5 Sv to 14.4 Sv (years 2005 to 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

We followed the RECCAP2-ocean protocol and divided the total sea-air CO\(_2\) flux (\(F^{\text{total}}\)) into five components. Specifically, the anthropogenic sea-air CO\(_2\) flux from increasing atmospheric CO\(_2\) in the atmosphere (\(F_{\text{ant}}\)) was divided into a steady-state component \(F_{\text{ant}^{\text{ss}}}\) representing the anthropogenic uptake 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 \(F_{\text{ant}}\). Like \(F_{\text{ant}}\), the natural sea-air flux of CO\(_2\) under pre-industrial atmospheric CO\(_2\) (\(F_{\text{nat}}\)) was divided into \(F_{\text{nat}^{\text{ss}}}\) under a constant climate (steady-state \(F_{\text{nat}^{\text{ss}}}\) or short \(F_{\text{nat}^{\text{ns}}}\), and the modulation of the \(F_{\text{nat}}\) due to climate variability and climate change (non-steady
state $F_{\text{nat}}$ or short $F_{\text{nat}}^{\text{ns}}$). The fifth flux component is the sea-air CO$_2$ flux due to the difference between the input of carbon and alkalinity across the land-sea interfaces from rivers and coastal erosion and the burial of carbon and alkalinity components in sediments ($F_{\text{nat}}^{\text{riv-bur}}$). 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 (RECCA P2)” 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 total flux across the sea-air interface ($F_{\text{total}}$) can thus be written as:

$$F_{\text{total}} = F_{\text{ant}}^{\text{ss}} + F_{\text{ant}}^{\text{ns}} + F_{\text{nat}}^{\text{ss}} + F_{\text{nat}}^{\text{ns}} + F_{\text{nat}}^{\text{riv-bur}}$$

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.

2.2.3 RECCAP2 simulations and their relation to CO$_2$ flux components

The RECCAP2 database 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. The four simulations all start in preindustrial times and extend through 2018, however, the GOBMs used different definitions of “preindustrial” with simulations starting between 1765 and 1870, and the corresponding assumed pre-industrial CO$_2$ mixing ratios varying between 278 ppm and 286 ppm. Simulations A and C were forced with historically increasing CO$_2$, whereas simulations B and D were forced with constant pre-industrial CO$_2$. 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). Afterwards, simulations A and D were forced with these historical atmospheric reanalysis fields, whereas simulations B and C continued with the same constant atmospheric
reanalysis fields that were applied before 1948 or 1958. Thus, each simulation represents a different combination of the CO$_2$ flux components:

- Simulation A is forced with historical atmospheric reanalysis data and historically increasing CO$_2$, yielding:
  \[ F_{\text{SimA}} \approx F_{\text{ant}}^{ss} + F_{\text{ant}}^{ns} + F_{\text{nat}}^{ss} + F_{\text{nat}}^{ns} + F_{\text{nat}}^{riv-bur}. \]  
  (2)

- Simulation B is forced with the same repeated annual atmospheric forcing and constant pre-industrial CO$_2$ levels, yielding:
  \[ F_{\text{SimB}} \approx F_{\text{nat}}^{ss} + F_{\text{nat}}^{riv-bur}. \]  
  (3)

- Simulation C is forced with a constant atmospheric forcing and historically increasing CO$_2$, yielding:
  \[ F_{\text{SimC}} \approx F_{\text{ant}}^{ss} + F_{\text{nat}}^{ss} + F_{\text{nat}}^{riv-bur}. \]  
  (4)

- Simulation D is forced with historical atmospheric reanalysis data and constant pre-industrial CO$_2$ levels, yielding:
  \[ F_{\text{SimD}} \approx F_{\text{nat}}^{ss} + F_{\text{nat}}^{ns} + F_{\text{nat}}^{riv-bur}. \]  
  (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.

The ocean physical and biogeochemical fields of the GOBMs were initialized with gridded observation-based estimates of ocean physics and biogeochemistry averaged over the last decades. The observation-based ocean DIC concentrations were thereby adjusted to represent pre-industrial DIC by removing the historical anthropogenic carbon uptake.

Four of the ten GOBMs considered here (FESOM-REcoM-LR, MOM6-Princeton, ORCA1-LIM3-PISCES, PlankTOM12) run the four simulations straight from these initial conditions without a pre-industrial spinup, while the remaining six (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 1586 years (overview table in DeVries et al. (in review)) using a repeated year of climatological atmospheric forcing and each model’s assumed pre-industrial atmospheric CO$_2$ with the goal to reach a near steady-state between the atmosphere and the ocean. Steady-state in this context refers to the state of a model under constant forcing, in which all multi-annual mean fluxes are time-invariant at the local scale and globally integrated zero. Few of the 6 models with spinup reach this objective, largely because of the spinup 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 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 PgC yr$^{-1}$, with a relatively small drift over time (Hauck et al., 2020). However, regionally, this effect can be more important. We call this bias in the sea-air CO$_2$ flux due to insufficient spinup and its drift
over time $F_{\text{drift+bias}}$. This $F_{\text{drift+bias}}$ 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 flux in one RECCAP2 simulation from the sea-air CO$_2$ flux in another RECCAP2 simulation, assuming that $F_{\text{nat-riv-bur}}$ and $F_{\text{drift+bias}}$ are not affected by increasing atmospheric CO$_2$ or changing atmospheric forcing across the varying simulations and that the different flux components add up to the total flux without substantial non-linearities:

$$F_{\text{nat-ss}} \approx F_{\text{SimC}} - F_{\text{SimB}}$$

(6)

$$F_{\text{nat-ns}} \approx F_{\text{SimA}} - F_{\text{SimC}} + F_{\text{SimB}} - F_{\text{SimD}}$$

(7)

$$F_{\text{nat-ns}} \approx F_{\text{SimD}} - F_{\text{SimB}}$$

(8)

The total air-sea CO$_2$ flux ($F_{\text{total}}$) can hence be estimated as follows:

$$F_{\text{total}} \approx F_{\text{SimA}} - F_{\text{SimB}} + F_{\text{nat-ss}} + F_{\text{nat-riv-bur}}$$

(9)

Globally, $F_{\text{nat-ss}}$ is by definition zero, so that only $F_{\text{nat-riv-bur}}$ has to be known for a GOBM-based estimate of $F_{\text{total}}$. Unfortunately, $F_{\text{nat-riv-bur}}$ cannot be quantified from the here-used GOBM simulations because their set-ups were not designed to represent riverine input and/or sediment burial (see Section 3.1.1). For the estimation of $F_{\text{total}}$ from GOBMs in RECCAP2, the observation-based estimate from (Regnier et al., 2022) was used in equation (9) as an approximation of global $F_{\text{nat-riv-bur}}$ (i.e., $0.65 \pm 0.30$ Pg C yr$^{-1}$), henceforth called $F_{\text{obs-riv-bur}}$. 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.

Regionally, estimating $F_{\text{total}}$ from these simulations is more difficult, because the regional $F_{\text{nat-ss}}$ is not zero as the ocean takes up and releases natural carbon regionally. Therefore, $F_{\text{total}}$ cannot be estimated as the difference between simulations A and B as this difference does not only remove $F_{\text{nat-drift+bias}}$ and $F_{\text{nat-riv-bur}}$, but also $F_{\text{nat-ss}}$. Hence, we estimate regional $F_{\text{total}}$ from simulation A and accept the simulated regional $F_{\text{nat-drift+bias}}$ and $F_{\text{nat-riv-bur}}$ as inherent uncertainties. To still estimate $F_{\text{total}}$, we added an observation-based estimate of the regional $F_{\text{nat-riv-bur}}$ ($F_{\text{obs-riv-bur}}$) to the sea-air CO$_2$ fluxes from simulation A. This regional observation-based estimate of $F_{\text{obs-riv-bur}}$ is derived from the estimated regional pattern of $F_{\text{nat-riv-bur}}$ (Lacroix et al., 2020), which is scaled with a constant factor for all grid cells such that the global integral matches the postulated global value of $F_{\text{obs-riv-bur}}$ of $0.65$ Pg C yr$^{-1}$. Overall, the impossibility to disentangle the regional values of $F_{\text{nat-ss}}$, $F_{\text{drift+bias}}$, and $F_{\text{nat-riv-bur}}$ in the models and the uncertainties of the regional observation-based $F_{\text{obs-riv-bur}}$ hence add additional uncertainty to the regional estimates of $F_{\text{total}}$. 

2.3 Observation-based estimates, their uncertainties and their usage for comparison with GOBMs

To compare the total sea-air CO$_2$ fluxes from the GOBMs with observation-based estimates, we utilize the RECCAP2 dataset of $p$CO$_2$ products, including AOML_EXTRAT, CMEMS-LSCE-FFNN, CSIR-ML6, JenaMLS, JMA-MLR, MPI-SOMFFN, OceanSODA-ETHZ, UOEX_Wat20, and NIES-MLR3 (see table in DeVries et al. (in review)) for references and further details). These $p$CO$_2$ 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 only, i.e., for the period when all products provide estimates.

The simulated regional $F_{\text{nat}}^{\text{ns}}$ were compared to ocean inversion-based estimates (Mikaloff Fletcher et al., 2007). These rely on observations of interior ocean DIC, alkalinity, and nutrients to create a conservative DIC tracer where the anthropogenic concentration in each grid cell is subtracted following Gruber et al. (1996) as well as changes in the ocean interior DIC due to biological processes. In a second step, 10 ocean circulation models were used to determine the circulation pattern by injecting a dye tracer at the ocean surface at a constant rate. Finally, the circulation pattern which results in the best fit with the observations of the adjusted DIC tracer is utilized to determine $F_{\text{nat}}^{\text{ss}}$.

To constrain the simulated $F_{\text{ant}}^{\text{ns+ss}}$, we used the mapped anthropogenic carbon storage between the years 1800 and 2002 from the GLODAPv2.2016b-product (Lauvset et al., 2016). This data-product is based on the TTD-method (Matear et al., 2003; Waugh et al., 2006) and henceforth referred to as TTD-estimate. It includes estimates of a mapping error, but a comprehensive error estimate containing observational, methodological, and mapping errors is not provided with the dataset. In lack of such an estimate, we utilized the error-estimate of ±29% for the $C_{\text{ant}}$-storage of the North Atlantic (Steinfeldt et al., 2009), which is a simplified and rather conservative error estimate (Khatiwala et al., 2013; Terhaar, Tanhua, et al., 2020). Additionally, the mapped $C_{\text{ant}}^{\text{ns+ss}}$-storage from the year 1800 to the year 1994 as well as that between the years 1994 and 2007 were quantified by Sabine et al. (2004) with the ocean tracer–based $\Delta C^*$ method (henceforth referred to as $\Delta C^*$-estimate) and by Gruber, Clement, et al. (2019) with the eMLR($C^*$)-method (henceforth referred to as eMLR($C^*$)-estimate), respectively. Uncertainties of the globally integrated estimates of both $C_{\text{ant}}^{\text{ns+ss}}$-storage estimates were provided when comparing these estimates to simulated values. We compared (changes of) anthropogenic carbon inventories between GOBMs and mapped TTD-, $\Delta C^*$- and eMLR($C^*$)-estimates (Section 3.3.1), respectively. As the mapped TTD-, $\Delta C^*$- and eMLR($C^*$)-estimates do not cover all Ocean basins (e.g., the Arctic Ocean and the Marginal Seas are not covered by the mapped TTD-estimate), the GOBM estimate is only integrated over grid-points that the associated mapped observation-based estimate covers. When referring to a comparison between TTD-, $\Delta C^*$- and eMLR($C^*$)-estimates and GOBM-estimate of interior ocean $C_{\text{ant}}$-storage then this excludes depth under 3000 m as well as the Arctic Ocean and the marginal Seas.

For the AMOC (here defined as maximum of the Atlantic meridional overturning streamfunction at 26°N), data from the RAPID-Meridional Overturning Circulation and Heatflux Array-Western Boundary Time Series array at 26°N were used (Frajka-Williams et al., 2021) to calculate the mean AMOC strength from 2005 to 2018. The measurement uncertainty of the annual AMOC values 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 (McCarthy et al., 2015)) and each year to be independent as the moorings of the observational array are exchanged every year.

The interfrontal sea surface salinity is the average sea surface salinity in the region limited by the polar front and the subtropical front and approximately describes the region where the upwelled circumpolar deep water is transformed into mode and intermediate water. Mean estimates and uncertainties were derived as described in Terhaar, Frölicher, et al. (2021) using gridded monthly climatologies of sea surface salinity and of sea surface temperature from the World Ocean Atlas 2018 (Locarnini et al., 2018; Zweng et al., 2018).

The volume of ventilated waters is defined as the volume of water south of 30°S with densities above the mean interfrontal sea surface density and below the mean interfrontal sea surface density plus 0.8 kg m\(^{-3}\). The value of 0.8 kg m\(^{-3}\) corresponds to approximately 2-3 times the area-weighted standard deviation of the monthly sea surface densities in the inter-frontal zone across the ensemble of ESMs used by Terhaar, Frölicher, et al. (2021). This density thus covers most of the denser water masses in the area that are relatively fast ventilated and excludes the small areas of very dense surface waters that very slowly ventilated a large amount of the deep ocean.

For comparisons of surface DIC and alkalinity between observation-based estimates and GOBMs, the observation-based 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) were used. As the gridded estimates of these three products are based on observations of surface ocean \(p\text{CO}_2\) and alkalinity in space and time, we henceforth call them \(p\text{CO}_2\)/alkalinity products. Furthermore, gridded GLODAPv2 estimates of the same variables were also used (Lauvset et al., 2016), where DIC is normalized to the atmospheric \(p\text{CO}_2\) of 2002. For comparison, output from the \(p\text{CO}_2\)/alkalinity products and GOBMs were averaged over the years 1986 to 2018, the longest time period available with the year 2002 in its center.

Additionally we compared the simulated and observation-based Revelle factor (Revelle and Suess, 1957), carbonate ion (\(\text{CO}_3^{2-}\)) concentrations, and the chemical surface ocean uptake capacity. \(\text{CO}_3^{2-}\) acts as a buffer for the ocean carbon uptake (Broecker et al., 1979), which declines with increasing \(\text{CO}_2\) uptake (Sarmiento & Gruber, 2006). The Revelle factor describes the overall uptake capacity of the ocean:

\[
\text{Revelle} = (\Delta \text{DIC} / \text{DIC}) / (\Delta [p\text{CO}_2] / [p\text{CO}_2]).
\]  

We re-arranged this equation to quantify the amount of additional carbon that the surface ocean can take up for a given increase in \(p\text{CO}_2 (\Delta \text{DIC} / \Delta [p\text{CO}_2])\) and defined this to be the chemical uptake capacity:

\[
\Delta \text{DIC} / \Delta [p\text{CO}_2] = \text{DIC} / (\text{Revelle} \times [p\text{CO}_2]).
\]

For consistency, the Revelle factor, \(\text{CO}_3^{2-}\), and the chemical uptake capacity were calculated based on the provided temperature, salinity, DIC, and alkalinity in GLODAPv2, the three \(p\text{CO}_2\)/alkalinity products, and all GOBMs using mocsy2.0 (Orr & Epitalon, 2015), respectively, 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).

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.,...
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, Kwiatkowski, et al., 2020; Terhaar, Frölicher, et al., 2021; Terhaar, Torres, et al., 2021; Terhaar et al., 2022).

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. When globally comparing the simulated $F_{\text{total}}$ of the GOBMs to that of the $pCO_2$ products, $F_{\text{obs}}^{riv-bur}$ has to be added to the GOBM estimate (see Section 2.2) and the relatively large 1-sigma uncertainty of $F_{\text{obs}}^{riv-bur}$ ($\pm 0.15 \text{ Pg C yr}^{-1}$) substantially increases the uncertainty of the GOBM-derived estimate. For the global $F_{\text{total}}$ estimates from GOBMs, we will therefore provide both a combined uncertainty (standard deviation of GOBM ensemble and of $F_{\text{obs}}^{riv-bur}$) and a pure standard deviation that does not include the uncertainty of $F_{\text{obs}}^{riv-bur}$ and hence is a measure of model-based differences only. Excluding the uncertainty of $F_{\text{obs}}^{riv-bur}$ allows comparing the ensemble spread of estimates of GOBMs to that of the $pCO_2$ products. Regionally, the uncertainty of $F_{\text{total}}$ is only provided as the standard deviation across the GOBM ensemble, because regional uncertainties of $F_{\text{obs}}^{riv-bur}$ 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). The RECCAP2 biome-mask is a slightly modified version of the oceanic biomes of Fay & 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 a late starting date

To quantify the difference in the simulated anthropogenic carbon uptake from 1980 to 2018 due to different starting dates (see Section 2.2.3), it would be ideal to re-run all simulations that started later than 1765 from 1765 onwards. However, spinning-up several GOBMs with another pre-industrial $pCO_2$ and re-running the historical simulations from 1765 to 2018 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 and one in 1850, 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 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 $F_{\text{total}}$ (-1.41±0.24 Pg C yr$^{-1}$; here excluding uncertainties of $F_{\text{obs riv-bur}}$) that is statistically indistinguishable from that estimated by the $pCO_2$-products (-1.71±0.26 Pg C yr$^{-1}$) (Table 1, Figure 1). In addition, the overall increasing trend is similarly represented by the two classes of estimates. Still, the difference of the long-term means of 0.30±35 Pg C yr$^{-1}$ (18±20% of the mean $pCO_2$-product estimate) is substantial. Moreover, the difference of annual mean fluxes between GOBMs and $pCO_2$-products varies with time, exceeding 20% of the average value of the $pCO_2$-products from 1985 to 1990, in 2009 and 2010, and from 2016 to 2018. 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 yr$^{-1}$ representing ~17% of their average CO$_2$-flux. Even larger differences are found on the regional scale (Figure 1b-f).

Table 1: Ensemble mean estimate of global and regional CO$_2$-fluxes (Pg C yr$^{-1}$) by GOBMs and $pCO_2$ products. The GOBM uncertainty excludes the uncertainty of $F_{\text{obs riv-bur}}$.

<table>
<thead>
<tr>
<th>Region</th>
<th>Global</th>
<th>Atlantic</th>
<th>Pacific</th>
<th>Indian</th>
<th>Arctic</th>
<th>Southern</th>
</tr>
</thead>
<tbody>
<tr>
<td>GOBMs</td>
<td>-1.41±0.24</td>
<td>-0.23±0.15</td>
<td>-0.34±0.12</td>
<td>-0.10±0.06</td>
<td>-0.06±0.03</td>
<td>-0.73±0.31</td>
</tr>
<tr>
<td>$pCO_2$ products</td>
<td>-1.71±0.26</td>
<td>-0.37±0.06</td>
<td>-0.39±0.14</td>
<td>-0.13±0.04</td>
<td>-0.08±0.05</td>
<td>-0.74±0.07</td>
</tr>
</tbody>
</table>

Regionally, the time-averaged $F_{\text{total}}$ 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 $F_{\text{total}}$ estimates in the Arctic Ocean also start to diverge (Figure 1e). Furthermore, the GOBMs and the $pCO_2$ products do not show the
same decadal variability of $F_{total}$ in the Southern Ocean (Figure 1f). The inter-model ensemble
spread of simulated $F_{total}$ is 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 offer 1985 to
2018). A separation of $F_{total}$ 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 spread of $F_{total}$ in GOBMs stems from $F_{ss}$ (Figure 2a, Table S1). Regionally, the spread of
$F_{total}$ is dominated by the spread of the sum of $F_{nat}^s$, $F_{nat}^{riv-bur}$, and $F_{drift+biass}$ in all basins but the
Arctic Ocean (Figure 2b-d, Table S1). The second largest contributions to the model spread are
$F_{ant}^{ss}$ and $F_{nat}^{ns}$. In the Arctic Ocean, the spread of the sum of $F_{nat}^s$, $F_{nat}^{riv-bur}$, and $F_{drift+biass}$ and the
spread of $F_{nat}^{ns}$ are of similar size (Figure 2e, Table 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,
while the model spread in $F_{nat}^{ns}$ is caused by the inter-model differences in simulated $p$CO$_2$ under
the melting sea ice (Yasunaka et al., in review).

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
$p$CO$_2$ products estimates, and propose adjustments to the GOBM results. A special focus will lie
on the Atlantic Ocean, where the long-term mean difference between GOBMs and $p$CO$_2$ 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 $p$CO$_2$ products is
largest.
Figure 1. Time series of global and regional sea-air CO$_2$ fluxes from 1980 to 2018 based on 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 pCO$_2$ products estimates (blue) for the a) 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.
Figure 2. Time series of sea-air CO$_2$ flux components globally and regionally from 1980 to 2018 based on GOBMs. The total sea-air CO$_2$ flux ($F_{\text{total}}$) integrated over each basin adjusted for the riverine-burial induced sea-air CO$_2$ flux (green) and the individual flux components from the GOBMs ($F_{\text{ant}}^{\text{ss}}$ in red, $F_{\text{ant}}^{\text{ns}}$ in orange, $F_{\text{nat}}^{\text{ns}}$ in purple, and the sum of $F_{\text{nat}}^{\text{ss}}$, $F_{\text{nat}}^{\text{riv-bur}}$ and $F_{\text{drift+bias}}$ 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 $F_{\text{total}}$ does not include the uncertainty of the riverine adjustment.
3.1 Sea-air CO₂ fluxes in the steady state control simulation

3.1.1 Carbon fluxes from rivers and into sediments

The input of riverine carbon $F_{\text{nat}}^{\text{riv}}$ and the sedimentation of carbon $F_{\text{nat}}^{\text{bur}}$ is treated in various ways across the ensemble of GOBMs and varies from 0.00 Pg C yr\(^{-1}\) to 0.61 PgC yr\(^{-1}\) and from 0.00 Pg C yr\(^{-1}\) to 0.74 Pg C yr\(^{-1}\), respectively (Table 2). The difference between $F_{\text{nat}}^{\text{riv}}$ and $F_{\text{nat}}^{\text{bur}}$ varies between -0.14 Pg C yr\(^{-1}\) and -0.54 Pg C yr\(^{-1}\) and is 0.10±0.23 Pg C yr\(^{-1}\) when averaged over the 8 GOBMs that provide all four simulations.

Table 2. Global ocean carbon fluxes (Pg C yr\(^{-1}\)) averaged from 1980 to 2018. Positive fluxes indicate fluxes out of the ocean, except for the land-sea river carbon fluxes. $F_{\text{nat}}^{\text{riv-bur}}$ was estimated as the difference between the land-sea river carbon flux and the burial in sediments, except for NorESM-OC1.2. $F_{\text{drift-bias}}$ was derived as the difference between $F_{\text{SimB}}$ and $F_{\text{nat}}^{\text{riv-bur}}$. The GOBM-ensemble values exclude MOM6-Princeton (see Section 2.1).

<table>
<thead>
<tr>
<th>Land-sea river carbon flux</th>
<th>Burial in sediments</th>
<th>$F_{\text{nat}}^{\text{riv-bur}}$</th>
<th>$F_{\text{SimB}}$</th>
<th>$F_{\text{drift-bias}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>CESM-ETHZ</td>
<td>0.33</td>
<td>0.25</td>
<td>0.08</td>
<td>0.00</td>
</tr>
<tr>
<td>CNRM-ESM2-1</td>
<td>0.61</td>
<td>0.74</td>
<td>-0.13</td>
<td>-0.14</td>
</tr>
<tr>
<td>EC-Earth3</td>
<td>0.61</td>
<td>0.47</td>
<td>0.14</td>
<td>0.25</td>
</tr>
<tr>
<td>FESOM-REcoM-LR</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>-0.35</td>
</tr>
<tr>
<td>MOM6-Princeton</td>
<td>0.18</td>
<td>0.10</td>
<td>0.08</td>
<td>-0.23</td>
</tr>
<tr>
<td>MRI-ESM2-0</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.17</td>
</tr>
<tr>
<td>NorESM-OC1.2</td>
<td>0.00</td>
<td>0.54</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>ORCA025-GEOMAR</td>
<td>0.00</td>
<td>0.34</td>
<td>-0.34</td>
<td>-0.36</td>
</tr>
<tr>
<td>ORCA1-LIM3-PISES</td>
<td>0.61</td>
<td>0.59</td>
<td>0.02</td>
<td>-0.26</td>
</tr>
<tr>
<td>GOBM-ensemble</td>
<td>0.27±0.30</td>
<td>0.37±0.27</td>
<td>-0.03±0.15</td>
<td>-0.09±0.23</td>
</tr>
</tbody>
</table>

To estimate $F_{\text{nat}}^{\text{riv-bur}}$, we use the first-order assumption that $F_{\text{nat}}^{\text{riv-bur}} = F_{\text{nat}}^{\text{riv}} - F_{\text{nat}}^{\text{bur}}$ for all GOBMs except NorESM-OC1.2 (Table 2). This assumption ignores the potential influence of alkalinity and nutrient fluxes from riverine and sedimentation (Gao et al., 2023; Terhaar, Orr, Ethé,
et al., 2019) as we only have this information for NorESM-OC1.2. In that model, the carbon burial flux that is larger than the 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, which 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 \(F_{\text{nat}}^{\text{riv-bur}}\) is almost zero (Table 2). With the adjusted \(F_{\text{nat}}^{\text{riv-bur}}\) for NorESM-OC1.2, the multi-model mean \(F_{\text{nat}}^{\text{riv-bur}}\) is \(-0.03\pm0.15\) Pg C yr\(^{-1}\). In comparison, the model spread associated with \(F_{\text{total}}\) is \(0.24\) Pg C yr\(^{-1}\). Although \(F_{\text{nat}}^{\text{riv-bur}}\) does not directly affect the global estimation of \(F_{\text{total}}\), it may substantially affect the regional estimates of \(F_{\text{total}}\) and \(F_{\text{ss}}\).

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 \(F_{\text{drift+bias}}\), quantified as the difference of \(F_{\text{SimB}}\) and \(F_{\text{nat}}^{\text{riv-bur}}\) (equation 3, Table 2), varies from \(-0.35\) to \(0.17\) Pg C yr\(^{-1}\), with an ensemble mean of \(-0.06\pm0.18\) Pg C yr\(^{-1}\). The model spread around \(F_{\text{drift+bias}}\) is of similar order as the model spread associated with the global \(F_{\text{total}}\) (0.24 Pg C yr\(^{-1}\)). We assume that this is mostly a consequence of a too short spinup and hence of models not being in a steady state, since 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 \(\pm0.002\) Pg C yr\(^{-2}\) for all GOBMs (Hauck et al., 2020). Although our estimation of \(F_{\text{drift+bias}}\) 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 sufficiently long spin-up in each model to reach steady state may thus narrow down inter-model differences of regional \(F_{\text{nat}}\) and \(F_{\text{total}}\).

3.1.3 Steady state natural sea-air CO\(_2\) flux

The mean \(F_{\text{SimB}}\) estimates of the GOBMs from 1980 to 2018 (Figure 2) are \(-0.11\pm0.14\) Pg C yr\(^{-1}\) for the Atlantic Ocean, \(0.21\pm0.13\) Pg C yr\(^{-1}\) for the Pacific Ocean, \(-0.06\pm0.06\) Pg C yr\(^{-1}\) for the Indian Ocean, and \(-0.06\pm0.01\) Pg C yr\(^{-1}\) for the Arctic Ocean. In the Southern Ocean, the \(F_{\text{SimB}}\) estimate of \(-0.04\pm0.27\) Pg C yr\(^{-1}\) of the GOBMs is twice as uncertain as in the other basins. The relatively large uncertainty in the Southern Ocean may partly be the result of large inter-model differences in the simulated \(F_{\text{nat}}\) fluxes, as dynamically complex regions like the Southern Ocean are difficult to simulate (Sallée et al., 2013). Inter-model differences in \(F_{\text{drift+bias}}\) likely also play a role for the uncertain \(F_{\text{SimB}}\) 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 \(F_{\text{drift+bias}}\) than in other ocean basins with less upwelling. The Southern hemisphere and especially the Southern Ocean are also the locations where the \(F_{\text{drift+bias}}\) tends to be largest in Earth System Models (Séférian et al., 2016).

When comparing \(F_{\text{SimB}}\) in the Southern Ocean to \(F_{\text{SimB}}\) globally, a significant relationship \((r^2 = 0.62, p = 0.01)\) with a slope of 1.03 can be identified (Figure 3a). This relationship suggests that global inter-model differences related to the sum of \(F_{\text{drift+bias}}\) \((-0.06\pm0.18\) Pg C yr\(^{-1}\)) and \(F_{\text{nat}}^{\text{riv-bur}}\) \((-0.03\pm0.15\) Pg C yr\(^{-1}\)) are indeed primarily stemming from the Southern Ocean, especially as
such a relationship occurs in no other ocean basin (Figure 3b and Figure S2). Based this assumption, we subtract the sum of global ocean $F_{\text{drift+bias}}$ and $F_{\text{nat riv-bur}}$ (-0.09±0.22 Pg C yr$^{-1}$ for the GOBM-ensemble without MOM6-Princeton) from the Southern Ocean $F_{\text{SimB}}$ for each GOBM separately. This adjustment leads to an estimate of Southern Ocean $F_{\text{nat ss}}$ of 0.05 ± 0.18 Pg C yr$^{-1}$, 0.10 Pg C yr$^{-1}$ larger and of opposite sign than the non-bias adjusted average $F_{\text{SimB}}$ across all GOBMs and with a 33% smaller spread. The major part (>80%) of this adjustment is due to $F_{\text{drift+bias}}$. In the other basins, the regional $F_{\text{SimB}}$ does not seem to be significantly impacted by the sum of $F_{\text{drift+bias}}$ and $F_{\text{nat riv-bur}}$ across the GOBM ensemble or these fluxes cancel each other out. In these basins, we assume $F_{\text{SimB}}$ to be approximately equal to $F_{\text{nat ss}}$.

![Figure 3. Relationship between global and regional sea-air CO$_2$ fluxes of simulation B for 9 GOBMs. The relationship between sea-air CO$_2$ fluxes (averaged for 1980-2018, negative: into the ocean) of the global ocean and a) Southern Ocean and b) Atlantic Ocean is shown. Represented is the natural sea-air CO$_2$ flux plus a potential sea-air CO$_2$ flux bias due to an interior ocean drift and a sea-air CO$_2$ flux related to carbon fluxes from rivers and into sediments (simulation B). The dashed line indicates a linear fit and the shading the projection uncertainty with a 68% uncertainty interval. The same relationship for the other ocean basins is shown in Figure S2.](image)

Our estimates of $F_{\text{nat ss}}$ can be compared to inverse estimates of $F_{\text{nat ss}}$ (Mikaloff Fletcher et al., 2007) (see also Section 2.3). These inverse estimates of $F_{\text{nat ss}}$ show larger uptake in the Atlantic (-0.24±0.08 Pg C yr$^{-1}$) and Pacific Ocean (-0.07±0.14 Pg C yr$^{-1}$), more outgassing in the Southern Ocean (0.44±0.11 Pg C yr$^{-1}$), and similar uptake in the Arctic (-0.02±0.01 Pg C yr$^{-1}$) and Indian Ocean (-0.12±0.04 Pg C yr$^{-1}$). 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). When changing the northern boundary of the Southern Ocean to 44°S, the adjusted regional $F_{\text{nat ss}}$ of the GOBMs changes to 0.27±0.19 Pg C yr$^{-1}$, still 0.18 Pg C yr$^{-1}$ smaller than the mean inverse-based estimate but within its uncertainties. Without the adjustment
for $F_{\text{drift+bias}}$ and $F_{\text{nat-burst}}$, the difference between the simulated and inverse-based estimate of $F_{\text{nat-ss}}$ in the Southern Ocean are larger.

3.2 Non-steady state natural sea-air CO$_2$ flux

Averaged between 1980 and 2018, the GOBMs simulate an outgassing global $F_{\text{nat-ss}}$ of 0.05±0.05 Pg C yr$^{-1}$. Here, we 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 $F_{\text{nat-ss}}$ increase from 1980 to 2018 at a rate of 0.07±0.02 Pg C yr$^{-1}$ decade$^{-1}$ (Figure 2, Figure 4a). The tropical Pacific and the Indian section of the Southern Ocean are the main contributors to the trend towards stronger $F_{\text{nat-ss}}$ carbon outgassing (Figure 4a, Figure 2). The average trend towards stronger outgassing of $F_{\text{nat-ss}}$ is to a small part compensated by a trend towards non-steady uptake of natural CO$_2$ in the Northern Pacific and the Arctic Ocean (Figure 2a; Figure 2e; Yasunaka et al. (in review)). Across the model ensemble, large inter-model differences in the mean $F_{\text{nat-ss}}$ 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 $F_{\text{nat-ss}}$ of 0.16±0.03 Pg C yr$^{-1}$ 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 $F_{\text{nat-ss}}$ 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 $F_{\text{nat-ss}}$ is the dominant contributor to the inter-annual and decadal variability of $F_{\text{total}}$ in GOBMs and is globally 6 times larger than the variability in the climate-driven variability in the anthropogenic sea-air CO$_2$ fluxes ($F_{\text{ant-ss}}$) and regionally 2 to 6 times larger (Figure 2). The simulated temporal variability of $F_{\text{total}}$ in the Pacific Ocean is driven by $F_{\text{nat-ss}}$ (Figure 2c) and resembles the variability of $F_{\text{total}}$ in the $p$CO$_2$ 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 \( C_{\text{ant}} \)-storage since the beginning of the industrial period simulated by the here analyzed 8 GOBM ensemble resembles that of the TTD- and \( \Delta C^* \)-estimate (Figure 5, Figure S3) and 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). While the TTD- and \( \Delta C^* \)-based estimates and the here analyzed 8 GOBMs agree that the largest accumulation of \( C_{\text{ant}} \) per surface area is 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 \( C_{\text{ant}} \)
from surface to 3000 m depth that accumulated since preindustrial times. The simulated multi-
model mean GOBM interior ocean $C_{ant}$ is 83±15 Pg C in 1994, 22% (23 Pg C) lower than the
$\Delta C^*$-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 from 1765 and 1870 (see
Section 3.4.1) and biases in GOBM dynamics and biogeochemistry (see Section 3.4.5). 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 $\Delta C^*$-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). Over this recent period, the GOBM 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($C^*$)-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 late 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 (Perez et al., to be submitted) and Southern Ocean (Hauck et al., to be submitted).
The eMLR($C^*$)-estimate indicates an anomalously high rate of $C_{ant}$-change in the South Atlantic
for the period 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 et al., to be submitted). 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 $F_{ant^*}$ 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) suggesting a bias in the model mean state that persists over
centuries. In the following sections, we will analyze the model set-ups, and simulated circulation
and biogeochemistry to identify reasons for the underestimation of $F_{ant^*}$ by the GOBM ensemble.
Figure 5: Column inventories of historic and contemporary anthropogenic carbon storage changes, integrated from surface to 3000m depth. Visualised are a) e) i) observation-based estimates and related model-estimates based on 8 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) b) c) d) show results for $C_{\text{ant}}^{\text{NS+SS}}$ from the $\Delta C^*$-estimate for the period 1800-1994 and GOBM estimates from start date of each simulation to 1994, d) e) f) g) show results for $C_{\text{ant}}^{\text{SS}}$ from the TTD-estimate for the period 1800-2002 and GOBM estimates from start date of each simulation to 2002, while panels i) j) k) l) show results for $C_{\text{ant}}^{\text{NS+SS}}$ 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.

### 3.3.2 Influence of pre-industrial atmospheric CO$_2$ mixing ratio on anthropogenic carbon uptake

The difference in the simulated sea-air CO$_2$ flux from 1980 to 2018 between the simulations starting in 1765 and those starting in 1850 is simulated by the EMIC Bern3D-LPX to be 0.04-0.06 Pg C yr$^{-1}$, depending on the ocean mixing strength (see Section 2.6 for details of this set-up). Regionally, most differences occur in regions of strong upwelling, such as the Southern Ocean (Figure 6b). 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 and more than 50% of this difference (9.8-13.7 Pg C) occurs after 1850.
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 in 1850. Time series of c) the anthropogenic sea-air CO$_2$ flux the from simulations starting in 1850 with weak (blue), medium (orange), and strong (green) ocean mixing, and time series d) of the difference in the anthropogenic sea-air CO$_2$ flux the between simulations starting in 1850 and 1765 for the same ocean mixing strengths.

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. 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 (40x41 horizontal cells and only 3 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 adjustment term for these water masses.

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 yr$^{-1}$ 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 CO₂ flux \( (F_{\text{ant}^{\text{ss}}}) \) 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 \( F_{\text{ant}^{\text{ss}}} \) and \( C_{\text{ant}} \) in both North Atlantic and Southern Ocean (Figure 5, Section 3.3.1), similar to previous GOBMs (Orr et al., 2001) and ESMs (Frölicher et al., 2015; Goris et al., 2018; Terhaar, Frölicher, et al., 2021).

![Image of simulated mean and intermodel spread of the steady-state anthropogenic CO₂ flux.](image)

**Figure 7: Simulated mean and intermodel spread of the steady-state anthropogenic CO₂ flux.**
Maps of (a) the multi-model mean and (b) multi-model standard deviation of the steady state anthropogenic sea-air CO₂ flux averaged from 1980 to 2018 for 8 GOBMs.

3.3.3.1 Role of ocean circulation on steady state anthropogenic sea-air CO₂-fluxes in the Atlantic and the Southern Ocean

In the Atlantic Ocean, the AMOC is the underlying driver of the uptake and storage of \( C_{\text{ant}} \). It transports surface waters with high \( C_{\text{ant}} \) (Pérez et al., 2013) and subsurface waters with low \( C_{\text{ant}} \) (Ridge & McKinley, 2020) northwards. The subsurface waters outcrop in the subpolar gyre and are hence a sink of \( C_{\text{ant}} \) (Ridge & McKinley, 2020). Both water masses are eventually transformed into deep water and transported southward. The AMOC is also the main driver of \( F_{\text{ant}^{\text{ss}}} \) differences in the Atlantic across ensembles of ESMs from CMIP5 and CMIP6 (Goris et al., 2018, 2023; Terhaar et al., 2022), linking \( F_{\text{ant}^{\text{ss}}} \) and the amount of \( C_{\text{ant}} \) that was transported below 1000 m across these model ensembles (Goris et al., 2018, 2023).

Correlations between \( F_{\text{ant}^{\text{ss}}} \) and (i) the AMOC at 26.5°N or (ii) the storage of \( C_{\text{ant}} \) between 1000 m and 3000 m in the high latitude North Atlantic also occur across this ensemble of GOBMs and can be used to identify emergent constraints (Figure 8a,b). In combination with the respective observation-based estimates, the average annual Atlantic \( F_{\text{ant}^{\text{ss}}} \) from 1980 to 2018 can be constrained from -0.39 ± 0.05 Pg C yr⁻¹ to -0.43 ± 0.06 Pg C yr⁻¹ when using the \( C_{\text{ant}^{\text{ss}}} \) storage and to -0.42 ± 0.05 Pg C yr⁻¹ when using the AMOC. The constraints identify a common bias in the GOBMs towards too small AMOC strengths (mean underestimation of 18%) and \( C_{\text{ant}^{\text{ss}}} \) storage below 1000m (mean underestimation of 22%), and hence Atlantic \( F_{\text{ant}^{\text{ss}}} \) (mean underestimation of
8-10%, depending on the used constraint). Nevertheless, the uncertainties around the Atlantic $F_{\text{ant}^{ss}}$ estimate cannot be reduced due to the relatively large uncertainty of the observation-based estimate in case of the $C_{\text{ant}^{ss}}$ storage as well as the relatively weak but significant correlation between the AMOC and the Atlantic $F_{\text{ant}^{ss}}$ ($r^2 = 0.54$, $p = 0.04$). 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 $F_{\text{ant}^{ss}}$ than expected from the fit, respectively (Figure 8b). The correlation of the $C_{\text{ant}^{ss}}$ storage and $F_{\text{ant}^{ss}}$ is stronger ($r^2 = 0.84$, $p = 0.001$) because both variables are more directly related to each other and coherently affected by the late starting date. The relationships between Atlantic $F_{\text{ant}^{ss}}$ and (i) AMOC and (ii) $C_{\text{ant}^{ss}}$ storage between 1000 m and 3000 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 $C_{\text{ant}^{ss}}$ storage).

In the Southern Ocean, the magnitude of $F_{\text{ant}^{ss}}$ 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, i.e., 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 $F_{\text{ant}^{ss}}$ uptake (Terhaar, Frölicher, et al., 2021). As the density in the region of interest is almost entirely driven by the salinity (Supplement of Terhaar, Frölicher, et al. (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 $F_{\text{ant}^{ss}}$ and (i) the inter-frontal sea surface salinity, i.e., the mean surface salinity in the subtropical-polar frontal zone ($r^2 = 0.57$, $p = 0.03$), and (ii) the volume of ventilated waters ($r^2 = 0.63$, $p = 0.03$) (Figure 8c,d).

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, et al., 2021). Moreover, the biomes are partly defined based on the location of these fronts, so that 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 et al., to be submitted). The constraint with the sea surface salinity as predictor reduces the magnitude of $F_{\text{ant}^{ss}}$ in the Southern Ocean slightly from -0.74±0.09 Pg C yr$^{-1}$ to -0.72±0.08 Pg C yr$^{-1}$ (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, et al., 2021) between the sea surface salinity and $F_{\text{ant}^{ss}}$ can partly be explained by different starting dates as GOBMs with a late or early starting date have a smaller or larger absolute $F_{\text{ant}^{ss}}$ than expected from the linear fit between the mean surface salinity in the subtropical-polar frontal zone, respectively (Figure 8c).

A common starting date of 1765 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 $F_{\text{ant}^{ss}}$ because the scarcity of subsurface observations would have resulted in large uncertainties of the observational constraint.
While the here considered emergent constraints change the average annual \( F_{\text{ant}} \) from 1980 to 2018 in Atlantic and Southern Ocean only slightly, the influence of circulation biases on \( F_{\text{ant}} \) increases in magnitude with increasing atmospheric \( F_{\text{ant}} \). Therefore, the difference between constrained and unconstrained \( F_{\text{ant}} \) increases over time (Figure S6) and a GOBM ensemble with circulation biases will have smaller trends in \( F_{\text{ant}} \) and deviate from the true \( F_{\text{ant}} \) with time.

**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) the Atlantic and c) the Southern Ocean, plotted against a) the Atlantic steady-state anthropogenic carbon storage between 1000 m and 3000 m depth for the year 2002, b) the Atlantic Meridional Overturning Circulation at 26°N averaged from 2005 to 2018, c) the inter-frontal 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) across GOBMs (green dots). The colors of the dots indicate the pre-industrial atmospheric \( p\text{CO}_2 \) for each GOBM. Observation-based estimates and their uncertainties are marked with dashed black lines and black shaded areas (see Section 2.4 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).

### 3.3.5.2 Surface ocean carbonate chemistry

The \( p\text{CO}_2/\text{alkalinity} \) products suggest that the largest chemical surface ocean uptake capacity (defined here as \( \Delta \text{DIC} / \Delta[p\text{CO}_2] \), 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 Ocean.
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 9a). 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 $p\text{CO}_2$/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 $\text{CO}_3^{2-}$ concentrations (Figure 9f), a high difference in surface 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 ~2300-2425 mmol m$^{-3}$) than ESMs (range of 2225-2415 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).

For the GOBMs, their globally different chemical uptake capacities do not explain their global differences in $\text{F}_{\text{ant}}^{\text{ss}}$ (Figure 9e), although studies with ESMs found such a relationship (Terhaar et al., 2022). Possible reasons for no emerging relationship between $\text{F}_{\text{ant}}^{\text{ss}}$ and the chemical uptake capacity, $\text{CO}_3^{2-}$, or the alkalinity across the GOBM ensemble are differences in $\text{F}_{\text{ant}}^{\text{ss}}$ due to different starting dates of the simulations (Section 3.3.2) and ongoing $\text{F}^{\text{drift+biased}}$. If a GOBM has a large negative or positive $\text{F}^{\text{drift+biased}}$, its upwelling waters have too low or high DIC, too high or low $\text{CO}_3^{2-}$, and hence a chemical uptake capacity that is too high or low, respectively. With time, the additional surface ocean DIC from $\text{F}^{\text{drift+biased}}$ reduces the chemical uptake capacity so that it is effectively smaller than the one expected from the theoretical chemical uptake capacity. Thus, $\text{F}^{\text{drift+biased}}$ adds considerable noise so that a potential relationship between the chemical uptake capacity and $\text{F}_{\text{ant}}^{\text{ss}}$ may not be identifiable. When considering only the four GOBMs with a longer spin-up than 1000 years, a relationship indeed emerges (Figure 9e-h).
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 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) 8 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 to 2018) of the accumulated global anthropogenic sea-air CO$_2$ flux against the global mean area-weighted
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 for the variables on the respective x-axis.

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 yr$^{-1}$ (Figure 10). As for $F_{nat}^{ns}$, we separate $F_{ant}^{ns}$ into an interannual variability component and a long-term linear trend component. On average, 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-variability that is mainly located in the North Atlantic subpolar gyre and in the Southern Ocean (Figure 10c) and not in the Pacific Ocean as for $F_{nat}^{ns}$ (Figure 4b,d). The interannual-variability is similar across the entire model ensemble (Figure 10d).

Regionally, $F_{ant}^{ns}$ is substantially smaller than regional $F_{nat}^{ns}$ underlining the relatively minor importance of anthropogenic non-steady state fluxes compared to natural 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 with recent trends in climate as also discussed by Lovenduski et al. (2008) and Hauck et al. (to be submitted). 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 North Atlantic subpolar gyre, 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 AMOC decline with climate change 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 8 GOBMs. Maps of the a) the multi-model mean and b) the multi-model standard deviation of the linear trend in anthropogenic non-steady state sea-air CO$_2$ flux without the inter-annual variability (calculated by fitting a linear trend) averaged from 1980 to 2018, as well as maps of c) the multi-model mean and d) the multi-model standard deviation of the inter-annual variability (linear trend is removed).

4 Discussion and resulting recommendations

4.1 Spin-up and associated biases in the sea-air CO$_2$ flux

As not all GOBMs have been fully spun-up, globally integrated $F^{\text{drift+bias}}$ varies from -0.35 Pg C yr$^{-1}$ to 0.17 Pg C yr$^{-1}$ across the GOBM ensemble (-0.06±0.18 Pg C yr$^{-1}$ on average). $F^{\text{drift+bias}}$ does not directly affect our estimate of the global $F^{\text{total}}$ (based on equation (9)) as $F^{\text{drift+bias}}$ is removed when subtracting $F^{\text{SimB}}$ from $F^{\text{SimA}}$. In addition, other effects from a GOBM not being in steady-state owing to an insufficient spinup, such as biases in temperature, salinity, DIC, or alkalinity, and consequent biases in the circulation or chemical uptake capacity may still affect $F^{\text{total}}$. Regionally, $F^{\text{drift+bias}}$ directly affects $F^{\text{total}}$ because subtracting $F^{\text{SimB}}$ from $F^{\text{SimA}}$ removes not only $F^{\text{drift+bias}}$ but also $F^{\text{nat}}$, which is regionally not zero. To regionally estimate $F^{\text{total}}$ from a GOBM, one could hence rely either on $F^{\text{SimA}} - F^{\text{SimB}}$ and add an independent estimate of $F^{\text{nat}}$ (e.g., the inverse model estimate from Mikaloff Fletcher et al. (2007)), which comes with its own uncertainties, or rely on $F^{\text{SimA}}$ and treat the regional $F^{\text{drift+bias}}$ as an inherent uncertainty (as done here). Most of $F^{\text{drift+bias}}$ is likely located in the Southern Ocean and hence mostly affects the regional
estimate of \( F^{\text{total}} \) in that ocean region. When assuming that \( F^{\text{drift+bias}} \) is almost entirely located in the Southern Ocean (Section 3.3.1), \( F^{\text{drift+bias}} \) could offset the total flux there (-0.7-0.8 Pg C yr\(^{-1}\)) by up to 50% in individual models and would increase the multi-model mean by \( \sim13\% \).

A model-by-model analysis would be necessary to determine the extent of the spinup related bias and drift in each GOBM and the necessary length of the spinup for a GOBM to reach steady state. Depending on the difference between the model’s steady-state and the initialization, the necessary length of the spinup may vary between individual GOBMs (Gürses et al., 2023) (see also Figure 11). Such a model-by-model assessment of the necessary spinup length would include the assessment of different variables in different regions and depth-ranges and exceeds the scope of this study. A comparison between the number of simulated years before the start of the analysis period of each GOBM and the \( F^{\text{drift+bias}} \) (Figure 11) suggests that a short spin-up is often insufficient to reduce \( F^{\text{drift+bias}} \) (Griffies et al., 2016; Orr et al., 2017; Séférian et al., 2016). While a longer spinup increases the computational costs, it provides a relatively simple way to reduce the uncertainty of the simulated \( F^{\text{total}} \) 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 model’s physics, biology, and carbonate chemistry.

![Figure 11](image-url)

**Figure 11.** Estimated global sea-air CO\(_2\) bias fluxes related to the models not being in steady-state for 9 GOBMs 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-flux (\( F^{\text{bias}} \)) is determined as specified in Section 3.2, Table 2. The spin-up ORCA025-GEOMAR was branched from a previous spin-up from the same model but with a coarser resolution.

4.2 Riverine and sediment fluxes

The GOBMs differ strongly in their representation of the riverine and sediment fluxes of carbon, nutrients and alkalinity, ranging from models without such fluxes to models with that attempt to resolve these fluxes explicitly. According to our approximation, none of the GOBMs simulates a resulting riverine and sediment flux-driven \( F^{\text{nat riv-bur}} \) comparable to the observation-based \( F^{\text{obs riv-bur}} \) of -0.65±0.30 Pg C yr\(^{-1}\) (Regnier et al., 2022). The different representation of the
riverine and sediment fluxes in the GOBMs thus represent an important inherent uncertainty of the model-simulated regional sea-air CO$_2$ fluxes and the global natural sea-air CO$_2$ fluxes. Global GOBM-estimates of $F_{\text{total}}$ are however unaffected by $F_{\text{nat-bur}}$ if equation (9) is used as the simulated $F_{\text{nat-bur}}$ is removed when subtracting $F_{\text{SimB}}$ from $F_{\text{SimA}}$ and replaced by the observation-based estimate of riverine and sediment fluxes. Apart from riverine carbon and alkalinity fluxes, an inadequate representation of riverine nutrient fluxes can also affect all components of the sea-air CO$_2$ fluxes 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). However, estimates of the impact of changing riverine carbon, alkalinity and nutrient fluxes 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). More research and model development is urgently needed to better represent the riverine and sediment fluxes in GOBMs to allow for a less uncertain quantification of global and regional sea-air CO$_2$ fluxes. An accurate observation-based estimate of the global riverine and burial derived sea-air CO$_2$ flux is necessary to estimate global $F_{\text{total}}$ for GOBMs without these fluxes. Despite large efforts over the last years (Lacroix et al., 2021; Regnier et al., 2022; Resplandy et al., 2018), the most recent observation-based estimate of the global $F_{\text{obs-bur}}$ of $0.65\pm0.30$ Pg C yr$^{-1}$ (Regnier et al., 2022) still has large uncertainties (−45%) that even exceed the simulated inter-model standard deviation of $F_{\text{total}}$ before accounting for $F_{\text{obs-bur}}$ ($\pm0.24$ Pg C yr$^{-1}$).

Regionally, the uncertainties of $F_{\text{obs-bur}}$ are even larger than globally. Across RECCAP2 chapters, the local distribution of $F_{\text{obs-bur}}$ is derived from Lacroix et al. (2021) (see Section 2.3.3), suggesting a strong riverine-burial-induced carbon outgassing in the Atlantic Ocean (0.27 Pg C yr$^{-1}$) and a relatively weak riverine-burial-induced carbon outgassing in the Southern Ocean (0.04 Pg C yr$^{-1}$). Contrarily, an older estimate by Aumont et al. (2001) suggests a smaller $F_{\text{obs-bur}}$ in the Atlantic Ocean and a larger $F_{\text{obs-bur}}$ in the Southern Ocean. One reason for this difference might be that Lacroix et al. (2021) quantify $F_{\text{obs-bur}}$ as the difference between a simulation with observation-based riverine fluxes of carbon 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 removal of the artificial surface ocean carbon and nutrient input may override the riverine signal, especially in regions far away from river deltas such as the Southern Ocean. The artificial carbon input in the reference simulation would also explain why the global estimate of $F_{\text{obs-bur}}$ of Lacroix et al. (2021) is smaller than other existing estimates (Aumont et al., 2001; Regnier et al., 2022; Resplandy et al., 2018). Another reason for the difference of the spatial distribution of $F_{\text{obs-bur}}$ between Lacroix et al. (2021) and Aumont et al. (2001) is the assumption of the lability of riverine organic matter, which is lower in Lacroix et al. (2021). 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. (2021) would instead occur in the Southern Ocean, $F_{\text{total}}$ in the Atlantic Ocean would double. Hence, more refined estimates of the lability of organic matter and its effect on $F_{\text{obs-bur}}$ are crucial to better constrain the total sea-air CO$_2$ flux and regional anthropogenic carbon sink estimates.

4.3. Starting date and pre-industrial CO$_2$
The different pre-industrial atmospheric CO\textsubscript{2} in each GOBM introduces a difference in the simulated anthropogenic carbon flux between the GOBMs (Section 3.3.2). We compared two estimates for the impact of a later starting date on the anthropogenic carbon fluxes, which suggest that a later starting date leads to a global underestimation of 0.04-0.06 Pg C yr\textsuperscript{-1} (3-5\% of $F_{\text{total}}$) for the period 1980-2018. However, this underestimation of 0.04-0.06 Pg C yr\textsuperscript{-1} is highly uncertain and possibly underestimated by about 40\%.

To avoid the need of an estimate of the underestimation and the uncertainties that come with it, our recommendation would be to start all simulations in 1765 where atmospheric CO\textsubscript{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 $C_{\text{ant}}$ and $F_{\text{ant}}$ (e.g., Khatiwala et al. (2009, 2013), Matsumoto & Gruber (2005), and Mikaloff Fletcher et al. (2006)). While this necessitates to perform up to 85 more years per simulation, the cost of running GOBMs in hindcast mode is much smaller than the cost of fully-coupled Earth System Models and computational constraints should thus not represent a major bottleneck.

### 4.4 Circulation biases

Previously identified relationships in ESMs between the AMOC and the North Atlantic $F_{\text{ant}}^{ss}$ (Goris et al., 2018) and the inter-frontal sea surface salinity and the Southern Ocean $F_{\text{ant}}^{ss}$ (Terhaar, Frölicher, et al., 2021) could also be identified in this GOBM ensemble. Overall, the considered GOBMs underestimate the strength of the AMOC (on average by 3.1±5.2 Sv at 26.5°N) and hence $F_{\text{ant}}^{ss}$ in the Atlantic and slightly overestimate the inter-frontal sea surface salinity (on average by 0.05±0.12) and hence $F_{\text{ant}}^{ss}$ in the Southern Ocean, though the resulting improvements of both constrained $F_{\text{ant}}^{ss}$ estimates are small for the ensemble average.

The on average relatively good agreement of the simulated and observed sea surface salinity between the polar and subtropical fronts in the Southern Ocean is a direct consequence of the forcing with atmospheric observation-based temperatures from reanalysis products such that the location of the fronts is well presented by the models. In addition, some of the GOBMs also restore the salinity at the ocean surface towards observed salinities. Despite this, some GOBMs still overestimate the salinity substantially.

The AMOC strength at 26°N, however, differs significantly across our considered GOBMs with its multi-model mean being negatively biased. 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, only CESM-ETHZ has an extraordinarily small AMOC, which was improved in a later simulation set-up version. This led to larger $F_{\text{ant}}^{ss}$ uptake in the Atlantic (see Figure 8b). The substantial change in the AMOC from 3.5 to 14.8 Sv in CESM-ETHZ due to a different sea surface salinity restoring timescale, i.e., a different artificial salinity flux across the air-sea interface, highlights the strong sensitivity of the ocean circulation to atmospheric fluxes.

The here used emergent constraints provide relatively robust relationships between circulation features and carbon fluxes, which were tested across the CMIP5 and CMIP6 ensembles and the here used GOBM ensemble. 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). While the best estimates of $F_{\text{ant}}^{ss}$ in the Atlantic and Southern Oceans have changed the original estimate by less than 10% here, other model ensembles might have larger biases and changes in $F_{\text{ant}}^{ss}$ might hence be different. The relatively small reduced uncertainty in both regions (<11%) is likely due to weaker correlations due to different pre-industrial $pCO_2$ in the ensemble of GOBMs, which can relatively easily be improved following our recommendation of starting all simulations in 1765. In the long-term, we recommend improving the representations of key ocean circulation metrics in the GOBMs.

4.5 Ocean biogeochemistry

The globally averaged chemical uptake capacity does not show a strong relationship with globally integrated $F_{\text{ant}}^{ss}$ across the GOBM ensemble (Figure 9e) although such a relationship was found across an ensemble of ESMs (Terhaar et al., 2022). Here, the relationship might be blurred by other processes that are influencing the simulated $F_{\text{ant}}^{ss}$, namely circulation biases, different starting dates and bias due to different spin-up length. Accounting for the influence of the bias in circulation on $F_{\text{ant}}^{ss}$ (Section 3.4.5.1), i.e., increasing $F_{\text{ant}}^{ss}$ in the North Atlantic for GOBMs with a too small AMOC, supports the relationship, but does not lead to a tighter relationship. If only GOBMs with a spin-up above 1000 years were considered, a linear relationship between $F_{\text{ant}}^{ss}$ and the chemical uptake capacity emerges (Figure 9e) that resembles the same relationship across ESMs (Terhaar et al., 2022). However, the small number of considered GOBMs and the range of observation-based estimates of the chemical uptake capacity does not allow to exploit such a potential relationship yet. Eventually, only a GOBM ensemble with all models being spun-up to steady-state and better constrained observation-based estimates would allow drawing such conclusions more robustly.

4.6 Gap between observation-based estimates and GOBMs

For the period 1985-2018, our analysis identifies a gap in $F_{\text{total}}$ of 0.30 Pg C yr$^{-1}$ between surface $pCO_2$ products (-1.71±0.26 Pg C yr$^{-1}$) and GOBMs (-1.41±0.28 Pg C yr$^{-1}$; uncertainty includes the ±0.15 Pg C yr$^{-1}$ 1-sigma uncertainty of the $F_{\text{obs}}^{riv-bur}$ estimate). The GOBM underestimation of 0.30±0.38 Pg C yr$^{-1}$ (~18% of the $F_{\text{total}}$ of the $pCO_2$ products) can partially be explained by the late starting date of the GOBM simulations, circulation biases, and potential biogeochemical biases in the GOBMs. In addition, comparisons of the simulated $C_{\text{ant}}$ storage since the beginning of the industrialization and over recent years from 1994 to 2007 to observation-based estimates also suggests that the GOBMs underestimate $F_{\text{ant}}^{ss}$ by 20-30%. Apart from our identified average gap between GOBM and $pCO_2$ product estimates of the ocean carbon sink, we confirm that the trends in the ocean sink since 2000 also differ globally and regionally (Friedlingstein et al., 2022; Hauck et al., 2020). Although these different trends suggest a divergence between GOBM estimates and $pCO_2$ products in recent years (Figure 1a), an increase in $F_{\text{total}}$ by around ~20% in each year accounting for an underestimation of the anthropogenic steady-state flux would change this perception. The difference in $F_{\text{total}}$ would not appear as a divergence of both estimates since 2000 but as a change from an underestimation of $F_{\text{total}}$ by the
$pCO_2$ products to an overestimation. Nevertheless, the growth rates of $F^{\text{total}}$ are different between GOBMs and $pCO_2$ products and uncertainties remain of how the ocean sink evolves.

Regionally, different trends in $F^{\text{total}}$ between GOBMs and $pCO_2$ products seem to be driven by a mismatch in the temporal evolution of the Southern Ocean carbon sink, and an increasing gap between both estimates in the Atlantic (Figure 1). 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, et al., 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, it remains an open question if differences between both estimates are due to the erroneous models or the extrapolation of sparse observations with temporal aliasing.

The increasing gap in the Atlantic after 2000, however, appears to result from a smaller $F^{\text{total}}$ trend in GOBMs than in $pCO_2$ products. This smaller-than-observed trend in GOBMs 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 show a biased-small AMOC, whose influence on $F_{\text{ant}}^{\text{ss}}$ increases with increasing atmospheric CO$_2$ (Section 3.3.3.1; Figure S6). Perez et al. (to be submitted) show that the disagreement in Atlantic $F^{\text{total}}$ trends between GOBMs and $pCO_2$ products is especially large in the subpolar North Atlantic. This relates well to our finding about AMOC-biases as the influence of AMOC-biases on $F_{\text{ant}}^{\text{ss}}$ is potentially highest in the subpolar gyre where subsurface waters low in $C_{\text{ant}}^{\text{outcrop}}$. Furthermore, a study with ESMs has shown that AMOC-biases are strongly correlated to SST-biases in the North Atlantic (Wang et al., 2014). While we did not analyze SST biases in the North Atlantic, Rodgers et al. (in review) 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 $F^{\text{total}}$ of the GOBMs being more sensitive to warming (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.

4.7 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 $F^{\text{total}}$ results from $F_{\text{nat}}^{\text{ns}}$, i.e., the sea-air flux of natural carbon due to climate variability and climate change. Globally, $F_{\text{nat}}^{\text{ss}}$ is also an important flux component as it allows comparing the estimated ocean carbon sink from surface ocean $pCO_2$ products, which quantify $F_{\text{ant}}^{\text{ss}}, F_{\text{ant}}^{\text{ns}}, F_{\text{nat}}^{\text{ns}},$ and $F_{\text{obs}}^{\text{riv-bur}}$ (Friedlingstein et al., 2022) to observation-based estimates of the interior ocean change of $C_{\text{ant}}$ (Gruber, Clement, et al., 2019), which quantifies only changes in $F_{\text{ant}}^{\text{ss}}$, and $F_{\text{ant}}^{\text{ns}}$. 
Previous approximations estimated the global $F_\text{nat}^{\text{ns}}$ from 1994 to 2007 to be $5\pm3$ Pg C (Gruber, Clement, et al. (2019); based on observation-based estimates of anthropogenic carbon fluxes storage changes and surface ocean fluxes), to be 1.3 Pg C (Friedlingstein et al. (2022), based on GOBMs) and to be 1.6±0.8 Pg C (Terhaar et al. (2022), based on ESM simulations). The GOBMs here estimate a $F_\text{nat}^{\text{ns}}$ 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 $p$CO$_2$ products in the Pacific Ocean, where most of the inter-annual variability is located, and differs in the Southern Ocean, where $p$CO$_2$ products suggest a strong decadal variability before 2000 and a different trend after 2000 (Gloege et al., 2021; Gruber, Landschützer, et al., 2019; Landschützer et al., 2015) (Figure 1). However, the sparse observations in the Southern Ocean pose a challenge for the observation-based estimates. For example, Gloege et al. (2021) showed that the 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 these differences in variability between between GOBMs and $p$CO$_2$ products in the Southern Ocean 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 $F_\text{nat}^{\text{ns}}$, the largest $F_\text{ant}^{\text{ns}}$ 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 $F_\text{ant}^{\text{ns}}$ in the Southern Ocean could be explained by additional free ocean surface due to climate change, which can thus take up more $C_\text{ant}$ or by more upwelling of old water with low $C_\text{ant}$ content (Le Quéré et al., 2007), which can also take up more $C_\text{ant}$. Both processes would lead to partial compensation by $F_\text{nat}^{\text{ns}}$ fluxes (Hauck et al., to be submitted; Lovenduski et al., 2008), with either more natural carbon being upwelled to the surface or more $C_\text{nat}$ being released with reduced ice cover.

### 4.8 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) and Hauck et al. (2020)). Hauck et al. (2020) found that GOBMs on average overestimate the observed $p$CO$_2$ 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 sink primarily because of circulation biases. The late-starting date and biases in the chemical uptake capacity in models also tend to enhance this underestimation. Fay & McKinley (2021) tested how well GOBMs resemble the $p$CO$_2$ 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 & McKinley (2021), such as the application of the local riverine adjustment by Lacroix et al. (2021), not accounting for each models’ simulated regional \( F_{nat\,riv\,-\,bur} \) 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 4.2) and our analysis suggests that \( F_{drift\,\,bias} \) and \( F_{nat\,riv\,-\,bur} \) are not evenly distributed. These adjustments affect the regional \( F_{total} \) and don’t allow for robust simulated estimates of the regional \( F_{total} \). Therefore, constraining the global \( F_{total} \) with regional \( F_{total} \) appears to be prone to large uncertainties and we recommend rather using underlying physical and biogeochemical processes for such constraints.

5 Conclusions

Our analysis of GOBMs helps to explain inter-model differences and differences between \( pCO_2 \)-products and ocean biogeochemistry models estimates of the ocean carbon sink (DeVries et al., in review; Friedlingstein et al., 2022). These differences can be divided into (i) differences in the simulation set-ups, i.e., starting year and model spin-up, (ii) dynamical differences, i.e., model physics and biogeochemistry, and (iii) differences in boundary fluxes across the land-sea and sea-sediment interfaces.

The differences in the simulation set-ups can be resolved relatively easily by (a) using the CO\(_2\) mixing ratio from 1765 as pre-industrial value and branching the historical simulation from the pre-industrial control simulation in 1765 and (b) increasing the spin-up period to reduce the uncertainty of the simulated \( F_{total} \) in relation to model drift and allows to pinpoint weaknesses of the GOBMs and relationships across the GOBMs which are more apparent in steady-state.

Although one might suspect that an increasing spin-up would cause models to diverge from observations, we have found no evidence for this in this GOBM ensemble (Figure 8 and 9). Starting simulations in 1765 is an attractive option as 85 years of simulation may remove a global bias that is at least 0.04-0.06 Pg C yr\(^{-1}\) in simulations that started in 1850 (underestimation of the sink). We here recommend using 1765 and not 1800 as in the TTD and \( \Delta C^- \) estimates as the difference between atmospheric pCO2 in 1765 and 1800 already has a substantial effect on the ocean carbon sink until today (Bronselaer et al., 2017). The bias due to a too short spin-up is already accounted for on a global level through subtraction of the flux of the control simulation and hence does not affect estimates of the global carbon sink, such as the Global Carbon Budget estimate (Friedlingstein et al., 2022). However, a too short spin-up does impact regional flux estimates, particularly in the Southern Ocean. Moreover, where the models not being in steady-state also influences 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.

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 while these presentations are not improved yet. Our results show that the 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 ($0.03\pm0.13$). Both 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 set-up 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 long-term, we recommend more complex adjustments within the set-ups of the GOBMs to reduce these biases.

The relatively poor representation of riverine and burial fluxes introduces another uncertainty to the simulated sea-air $CO_2$ fluxes. 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 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 total sea-air $CO_2$ fluxes. Improving the representation of these fluxes and their underlying processes is thus of importance to better understand the regional ocean carbon sink.

As 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), it remains challenging to compare the modeled estimates to the observation-based estimates of the ocean carbon sink. 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 $F_{total}$, or has to be subtracted from the $pCO_2$ products to be able to compare these estimates to 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) and Lacroix et al. (2020)), the regional distribution and temporal variability of these fluxes still remains highly uncertain and renders a comparison between simulated and observation-based estimates of the ocean carbon sink complicated.

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 (late starting date, circulation biases, biogeochemical biases, biases in $C_{ant}$ storage) suggest that the GOBMs underestimate the ocean carbon sink on average. 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. Although explanations exist for the difference in the long-term mean carbon sink, the difference between the growth rates of the ocean carbon sink since 2000 globally, and in the Southern and Atlantic Oceans remains unresolved.

Overall, the model evaluation has helped to give recommendations for the set-up 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 in 1765, and increase the spin-up 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 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.

Acknowledgments

The authors have no conflicts of interest. J. Terhaar acknowledges funding from the Woods Hole Oceanographic Institution Postdoctoral Scholar Program, and the Swiss National Science Foundation under grant #200020_200511. 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. 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 number VH-NG-1301). 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. J. Terhaar thanks the science team and the crew of the R/V Sikuliaq for their great kindness, professional working atmosphere, great food, and moral support during the cruise in the Arctic Ocean in November, which provided an excellent environment for advancing this work.

Open Research

All of the RECCAP2 data will be made available in a public repository before publication.

References


Global Biogeochemical Cycles

Supporting Information for

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

Jens Terhaar1,2,3,†, Nadine Goris4,†, Jens D. Müller5, Tim DeVries6,7, Nicolas Gruber5, Judith Hauck8, Fiz F. Perez9, Roland Séférian10

1Department of Marine Chemistry and Geochemistry, Woods Hole Oceanographic Institution, Woods Hole, 02543, Massachusetts, USA

2Climate and Environmental Physics, Physics Institute, University of Bern, Switzerland

3Oeschger Centre for Climate Change Research, University of Bern, Switzerland

4NORCE Climate & Environment, Bjerknes Centre for Climate Research, Bergen, Norway

5Environmental Physics, Institute of Biogeochemistry and Pollutant Dynamics, ETH Zurich, Zürich, Switzerland

6Department of Geography, University of California, Santa Barbara, Santa Barbara, CA, USA

7Earth Research Institute, University of California, Santa Barbara, Santa Barbara, CA, USA

8Alfred-Wegener-Institut, Helmholtz-Zentrum für Polar- und Meeresforschung, Am Handelshafen 12, 27570 Bremerhaven, Germany

9Instituto de Investigaciones Marinas (IIM), CSIC, Vigo, Spain. (0000-0003-4836-8974)

10CNRM (Université de Toulouse, Météo-France, CNRS), Toulouse, France

Contents of this file

Text S1
Figures S1 to S10
Table S1
The GOBM PlankTOM12 is not considered in this study for anthropogenic flux and storage analysis as Planktom12 simulates almost no accumulation of $C_{\text{Ant}}$ in the high latitude North Atlantic but a large accumulation in the high-latitude North Pacific, where neither the other GOBMs nor the observation-based estimate from GLODAPv2 indicate an accumulation of $C_{\text{Ant}}$. This anomalous pattern is driven by a strong negative upper ocean (0-150 m) salinity bias in the North Atlantic and a positive salinity bias in the North Pacific. These biases in the salinity cause upper ocean density biases (Figures S9 and S10), which cause the overly strong accumulation of $C_{\text{Ant}}$ in the North Pacific and the overly weak $C_{\text{Ant}}$ accumulation in the North Atlantic. The salinity-density bias also extends in the surface Southern Ocean, where the inter-frontal sea surface density of PlankTOM12 is the lowest density among all GOBMs (1026.91 kg m$^{-3}$). The salinity bias decreases over the time of the simulation period (1980 to 2018) to almost zero. The size and the temporal evolution of the bias suggests that the model has been far away from equilibrium at the start of the analyzing period, possibly due to a change in the external forcing or other model set-up changes.
Figure S1. Biome-mask as utilized in RECCAP2 and within this study. The shading of the biomes is based on the ocean basin considered (Atlantic: orange-red, Pacific: yellow-orange, Southern Ocean: shades of blue, Indian Ocean: shades of lilac, Arctic: shades of green). In most cases, the letters before the hyphen denote the geographical location of the biome (NA: North Atlantic, SA: South Atlantic, AEQ: Atlantic equator, NP: North Pacific, SP: South Pacific, PEQ: Pacific Equator, SO: Southern Ocean) and the following letters describe their characteristics (SPSS: subpolar and seasonally stratified, STSS: subtropical and seasonally stratified, STPS: subtropical and permanently stratified, ICE: marginal sea-ice and U: upwelling). The Pacific equatorial upwelling biome is divided in east (E) and west (W). The newly introduced biomes here only describe the ocean region by name and are mostly self-explanatory with two exceptions (MED: Mediterranean Sea, BoB: Bay of Bengal).
Figure S2. Regional ocean sea-air CO₂ fluxes against global sea-air CO₂ flux in simulation B for 10 GOBMs. For each GOBM the regional sea-air CO₂ flux in simulation B in the a) Pacific, b) Indian, and c) Arctic Ocean, representing the natural sea-air CO₂ flux plus a potential sea-air CO₂ flux due to an interior ocean drift or carbon fluxes from rivers and into sediments is plotted against the global ocean sea-air CO₂ flux in simulation B. The dashed line indicates a linear fit and the shading the projection uncertainty with a 68% uncertainty interval.
Figure S3: Estimated integrated anthropogenic carbon storage (steady state) from surface to 3000m depth and from initial year until 2002 for each of the considered models as well as the TTD-based estimate from GLODAPv2.2. GLODAPv2.2 assumes a pre-industrial atmospheric CO2 mixing ratio of 280 ppm, while the pre-industrial atmospheric CO2 mixing ratio varies between 278 and 287.4 for GOBMs (see Section 2.2.3).
Figure S4: Estimated integrated anthropogenic carbon storage from 1994 to 2007 (non-steady and steady state), integrated from surface to 3000 m. Panels a)-i) show the estimate for each of the considered models, while panel j) shows the observation-based estimate derived with the eMLR(C*)-method (Gruber et al., 2009).
Figure S5. Relationship between long-term and recent change in marine anthropogenic carbon. Anthropogenic carbon ($C_{\text{ant}}$) change from pre-industrial times until 1994 compared to the change from 1994 to 2007. Colored dots show results from GOBMs and the colors show atmospheric pCO2 in the pre-industrial control simulations. The dashed green line indicates the linear fit across the GOBM ensemble, and the green shading shows the 68% projection uncertainties. Observation-based estimates from Sabine et al. (2004) for $C_{\text{ant}}$ in 1994 and from Gruber et al. (2019) for the change in $C_{\text{ant}}$ from 1994 to 2007 and their respective uncertainties are shown as dashed black lines and black shading.
Figure S6. Temporal evolution of the constrained steady-state anthropogenic carbon uptake in the Atlantic and Southern Ocean. Emergent constraints between the simulated steady-state anthropogenic carbon uptake for the year 1980 (red dots) and 2015 (blue dots) of a) the Atlantic and c) the Southern Ocean, plotted against a) the Atlantic Meridional Overturning Circulation at 26°N averaged from 2005 to 2018, b) the inter-frontal sea surface salinity averaged from 1980 to 2018. Included are linear fits (red/blue dashed lines) with 68% projection intervals (red/blue shaded area) across GOBMs (red/blue dots) as well as observational estimates and their uncertainties (dashed black lines and black shaded area). The crosses in a) indicate an additional simulation with CESM-ETHZ (see Section 2.1 of the main article). Panels b) and d) illustrate the temporal evolution of the steady-state anthropogenic carbon uptake in b) Atlantic and d) Southern Ocean, featuring both the unconstrained estimate (gray line and gray shaded area) as well as the constrained estimate and its uncertainty (green line and green shaded area) when constraining each annual value individually using b) the Atlantic Meridional Overturning Circulation at 26°N averaged from 2005 to 2018, and d) the inter-frontal sea surface salinity averaged from 1980 to 2018. The dashed green line in b) illustrates the mean constrained estimate when using the Cant storage below 1000m as observational constraint.
Figure S7. Surface ocean chemical uptake capacity in individual GOBMs normalized to 2002. Maps of the average increase in DIC per increase in $p$CO$_2$ averaged from 1982 to 2018 based as simulated by a) CESM-ETHZ, b) CNRM-ESM2-1, c) EC-Earth3, d) FESOM REcoM LR, e) MRI-ESM2-1, f) NorESM-OC1.2, g) ORCA025-GEOMAR, and h) ORCA1-LIM3-PISCES.
Figure S8. Difference between surface ocean chemical uptake capacity normalized to 2002 in individual GOBMs and \( pCO_2/\text{alkalinity} \) products. Maps of the difference in average increase in DIC per increase in \( pCO_2 \) averaged from 1982 to 2018 based as simulated by a) CESM-ETHZ, b) CNRM-ESM2-1, c) EC-Earth3, d) FESOM ReCoM LR, e) MRI-ESM2-1, f) NorESM-OC1.2, g) ORCA025-GEOMAR, and h) ORCA1-LIM3-PISCES compared to the average of \( pCO_2/\text{alkalinity} \) products.
Figure S9: Neutral density zonal mean section for the Atlantic Ocean. Differences between individual GOBMs and estimates calculated from observations provided through GLODAPv2.2021 (Lauvset et al., 2021). GOBM and GLODAP data were merged on the 1°x1° horizontal grid of the models, and GOBM data were linearly interpolated in the vertical dimension to match the exact sampling depth of the observations. The zonal mean sections do not account for biases in the spatio-temporal distribution of observations.
Figure S10: Neutral density zonal mean section for the Pacific Ocean. Differences between individual GOBMs and estimates calculated from observations provided through GLODAPv2.2021 (Lauvset et al., 2021). GOBM and GLODAP data were merged on the 1°x1° horizontal grid of the models, and GOBM data were linearly interpolated in the vertical dimension to match the exact sampling depth of the observations. The zonal mean sections do not account for biases in the spatio-temporal distribution of observations.
Table S1. Averaged annual standard deviations across the entire model ensemble per flux component and ocean basin. Units are in Pg C yr$^{-1}$.

<table>
<thead>
<tr>
<th>Region</th>
<th>$F_{\text{total}}$</th>
<th>$F_{\text{ant ss}}$</th>
<th>$F_{\text{ant ns}}$</th>
<th>$F_{\text{nat ss}} + F_{\text{riv-bur}} + F_{\text{drift+bias}}$</th>
<th>$F_{\text{nat ns}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Global Ocean</td>
<td>0.24</td>
<td>0.21</td>
<td>0.04</td>
<td>/</td>
<td>0.11</td>
</tr>
<tr>
<td>Atlantic Ocean</td>
<td>0.14</td>
<td>0.06</td>
<td>0.01</td>
<td>0.13</td>
<td>0.02</td>
</tr>
<tr>
<td>Pacific Ocean</td>
<td>0.15</td>
<td>0.08</td>
<td>0.02</td>
<td>0.14</td>
<td>0.09</td>
</tr>
<tr>
<td>Indian Ocean</td>
<td>0.08</td>
<td>0.03</td>
<td>0.01</td>
<td>0.06</td>
<td>0.03</td>
</tr>
<tr>
<td>Arctic Ocean</td>
<td>0.03</td>
<td>0.00</td>
<td>0.00</td>
<td>0.02</td>
<td>0.02</td>
</tr>
<tr>
<td>Southern Ocean</td>
<td>0.29</td>
<td>0.09</td>
<td>0.03</td>
<td>0.26</td>
<td>0.09</td>
</tr>
</tbody>
</table>