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

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

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

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

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

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

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

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

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

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

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

[10]CNRM (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 CO_2-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 pCO_2 and interior ocean measurements. Here, we provide an in-depth evaluation of ocean carbon sink estimates from 1980 to 2018 from a GOBM ensemble. As sources of inter-model differences and ensemble-mean biases our study identifies 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 CO_2 concentrations is known as anthropogenic carbon (C_{ant}; Sarmiento et al., 1992), while the DIC that existed prior to the start of the industrial revolution is called natural carbon (C_{nat}). Second order terms in the historical evolution of the ocean carbon sink are climate-change and climate-variability driven changes in the anthropogenic and natural air-sea CO_2 fluxes (Joos et al., 1999; 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 C_{ant} flux is controlled by the rate of C_{ant} transport from the surface ocean to the deep ocean, which depends on the concentration of C_{ant} in the surface ocean (Broecker et al., 1979) and the surface-to-deep water volume transport (Caldeira & Duffy, 2000; Mikaloff Fletcher et al., 2006; Orr et al., 2001; Sarmiento et al., 1992). In contrast, the air-sea flux of C_{nat} is primarily controlled by the interaction of surface heating or cooling affecting the solubility of CO_2 in seawater and transport and mixing, and biological processes of photosynthesis, respiration, and $CaCO_3$ production (Sarmiento & Gruber, 2006). While there is agreement on these drivers for C_{ant} and C_{nat} fluxes and their relative importance, an accurate quantification of the carbon sink and its processes is still challenging.

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

The RECCAP2 project is a continuation of the large efforts that have been undertaken in the last decades to quantify the past and present ocean carbon sink with pCO_2 products (Chau et al., 2022; Gregor 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 pCO_2-products (1990s: -1.91±0.25 Pg C yr^{-1} in models vs -2.14±0.34 Pg C yr^{-1} for pCO_2 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 pCO_2 products in the 2010s is around half as large as the annual CO_2 emissions in the United States of America over

the same period (Friedlingstein et al., 2022). This highlights the need for a more rigorous quantification of the ocean carbon sink to fully close the global carbon budget (Hauck et al., 2020). A better understanding of the fidelity of GOBMs is also needed if such models are to be used for monitoring, reporting, and verification of ocean-based carbon dioxide removal techniques (Gattuso et al., 2018).

Prior GOBM intercomparison studies (Khatiwala et al., 2013; Orr et al., 2001; Wanninkhof et al., 2013) and studies with related Earth System Models (ESMs) suggest several reasons for the differences mentioned above. Among them are biases in model dynamics such 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 (Bronselaer 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 pCO_2 products at a global and regional level. We also investigate the underlying reasons for these differences and provide recommendations for future assessments of the ocean carbon sink using GOBMs.

2 Materials and Methods

2.1 Ocean biogeochemistry models

The GOBMs analyzed in this study are general ocean circulation models with coupled sea ice and ocean biogeochemistry model components. They simulate the transport of biogeochemical tracers through advection and mixing and 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^{total}) into five components. Specifically, the anthropogenic sea-air CO_2 flux from increasing atmospheric CO_2 in the atmosphere (F_{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_{ant}. Like F_{ant}, the natural sea-air flux of CO_2 under pre-industrial atmospheric CO_2 (F_{nat}) was divided into F_{nat} under a constant climate (steady-state F_{nat} or short $F_{\text{nat}}^{\text{ss}}$), and the modulation of the F_{nat} due to climate variability and climate change (non-steady

state F_{nat} or short F_{nat}^{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 ($F_{nat}^{riv\text{-}bur}$). 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 (RECCAP2)" consider $F_{nat}^{riv\text{-}bur}$ to be an integral part of F_{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^{total}) can thus be written as:

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

Throughout this paper, carbon inventories are referred to as "*C*" in analogy to the fluxes that are abbreviated with "*F*". The same indices as for the fluxes were used to distinguish the respective components of carbon inventories and their change over time.

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 simulations 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:

$$F^{\text{SimA}} \approx F_{\text{ant}}{}^{\text{ss}} + F_{\text{ant}}{}^{\text{ns}} + F_{\text{nat}}{}^{\text{ss}} + F_{\text{nat}}{}^{\text{ns}} + F_{\text{nat}}{}^{\text{riv-bur}}. \quad (2)$$

- Simulation B is forced with the same repeated annual atmospheric forcing and constant pre-industrial CO_2 levels, yielding:

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

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

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

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

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

Simulations with a constant atmospheric climate (B, C) represent steady-state processes, while simulations with a variable climate (A, D) represent both steady-state and non-steady state processes. Similarly, simulations with rising CO_2 (A, C) represent both natural and anthropogenic CO_2 fluxes, while simulations with pre-industrial CO_2 (B, D) represent only natural CO_2 fluxes.

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 $F^{drift+bias}$. This $F^{drift+bias}$ 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 $F_{nat}^{riv\text{-}bur}$ and $F^{drift+bias}$ 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:

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

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

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

The total air-sea CO_2 flux (F^{total}) can hence be estimated as follows:

$$F^{total} \approx F^{SimA} - F^{SimB} + F_{nat}^{ss} + F_{nat}^{riv\text{-}bur} \quad (9)$$

Globally, F_{nat}^{ss} is by definition zero, so that only $F_{nat}^{riv\text{-}bur}$ has to be known for a GOBM-based estimate of F^{total}. Unfortunately, $F_{nat}^{riv\text{-}bur}$ 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^{total} from GOBMs in RECCAP2, the observation-based estimate from (Regnier et al., 2022) was used in equation (9) as an approximation of global $F_{nat}^{riv\text{-}bur}$ (i.e., 0.65±0.30 Pg C yr^{-1}), henceforth called $F_{obs}^{riv\text{-}bur}$. 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^{total} from these simulations is more difficult, because the regional F_{nat}^{ss} is not zero as the ocean takes up and releases natural carbon regionally. Therefore, F^{total} cannot be estimated as the difference between simulations A and B as this difference does not only remove $F_{nat}^{drift+bias}$ and $F_{nat}^{riv\text{-}bur}$, but also F_{nat}^{ss}. Hence, we estimate regional F^{total} from simulation A and accept the simulated regional $F_{nat}^{drift+bias}$ and $F_{nat}^{riv\text{-}bur}$ as inherent uncertainties. To still estimate F^{total}, we added an observation-based estimate of the regional $F_{nat}^{riv\text{-}bur}$ ($F_{obs}^{riv\text{-}bur}$) to the sea-air CO_2 fluxes from simulation A. This regional observation-based estimate of $F_{obs}^{riv\text{-}bur}$ is derived from the estimated regional pattern of $F_{nat}^{riv\text{-}bur}$ (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 $F_{obs}^{riv\text{-}bur}$ of 0.65 Pg C yr^{-1}. Overall, the impossibility to disentangle the regional values of F_{nat}^{ss}, $F^{drift+bias}$, and $F_{nat}^{riv\text{-}bur}$ in the models and the uncertainties of the regional observation-based $F_{obs}^{riv\text{-}bur}$ hence add additional uncertainty to the regional estimates of F^{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 pCO_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 pCO_2 products are a product of statistical models and sparse observations of surface ocean partial pressure of CO_2. We calculate long-term averages and trends over these products for the period 1985 through 2018 only, i.e., for the period when all products provide estimates.

The simulated regional F_{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_{nat}^{ss}.

To constrain the simulated F_{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_{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_{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 ΔC^* method (henceforth referred to as ΔC^*-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_{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-, ΔC^*- and eMLR(C^*)-estimates (Section 3.3.1), respectively. As the mapped TTD-, ΔC^*- 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-, ΔC^*- and eMLR(C^*)-estimates and GOBM-estimate of interior ocean C_{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 pCO_2 and alkalinity in space and time, we henceforth call them pCO_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 pCO_2 of 2002. For comparison, output from the pCO_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:

$$\text{Revelle} = (\Delta\text{DIC} / \text{DIC}) / (\Delta[pCO_2] / [pCO_2]). \quad (10)$$

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

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

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 pCO_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.,

2019; Hall et al., 2019). To obtain the constrained variables and their uncertainties, we here followed the approach from Cox et al. (2013) that has been frequently used over the recent years in ocean biogeochemistry (Bourgeois et al., 2022; Goris et al., 2018, 2023; Kwiatkowski et al., 2017; Terhaar, 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, $F_{obs}^{riv\text{-}bur}$ has to be added to the GOBM estimate (see Section 2.2) and the relatively large 1-sigma uncertainty of $F_{obs}^{riv\text{-}bur}$ (± 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 $F_{obs}^{riv\text{-}bur}$) and a pure standard deviation that does not include the uncertainty of $F_{obs}^{riv\text{-}bur}$ and hence is a measure of model-based differences only. Excluding the uncertainty of $F_{obs}^{riv\text{-}bur}$ allows comparing the ensemble spread of estimates of GOBMs to that of the pCO_2 products. Regionally, the uncertainty of F^{total} is only provided as the standard deviation across the GOBM ensemble, because regional uncertainties of $F_{obs}^{riv\text{-}bur}$ are not quantified so far.

2.5 Definition of ocean basins and sub-basin biomes

For our analysis, we applied the RECCAP2 biome-mask and the associated definition of ocean basins (Figure S1). 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^{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 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 $F_{\text{obs}}^{\text{riv-bur}}$.

	Global	Atlantic	Pacific	Indian	Arctic	Southern
GOBMs	-1.41±0.24	-0.23±0.15	-0.34±0.12	-0.10±0.06	-0.06±0.03	-0.73±0.31
pCO$_2$ products	-1.71±0.26	-0.37±0.06	-0.39±0.14	-0.13±0.04	-0.08±0.05	-0.74±0.07

Regionally, the time-averaged F^{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^{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 offor 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_{ant}^{ss} (Figure 2a, Table S1). Regionally, the spread of F^{total} is dominated by the spread of the sum of F_{nat}^{ss}, $F_{nat}^{riv\text{-}bur}$, and $F^{drift+bias}$ 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}^{ss}, $F_{nat}^{riv\text{-}bur}$, and $F^{drift+bias}$ 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 pCO_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 pCO_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 pCO_2 products estimates is largest, and on the Southern Ocean, where the various GOBM estimates differ the most and where the decadal variability of the difference between GOBMs and pCO_2 products is largest.

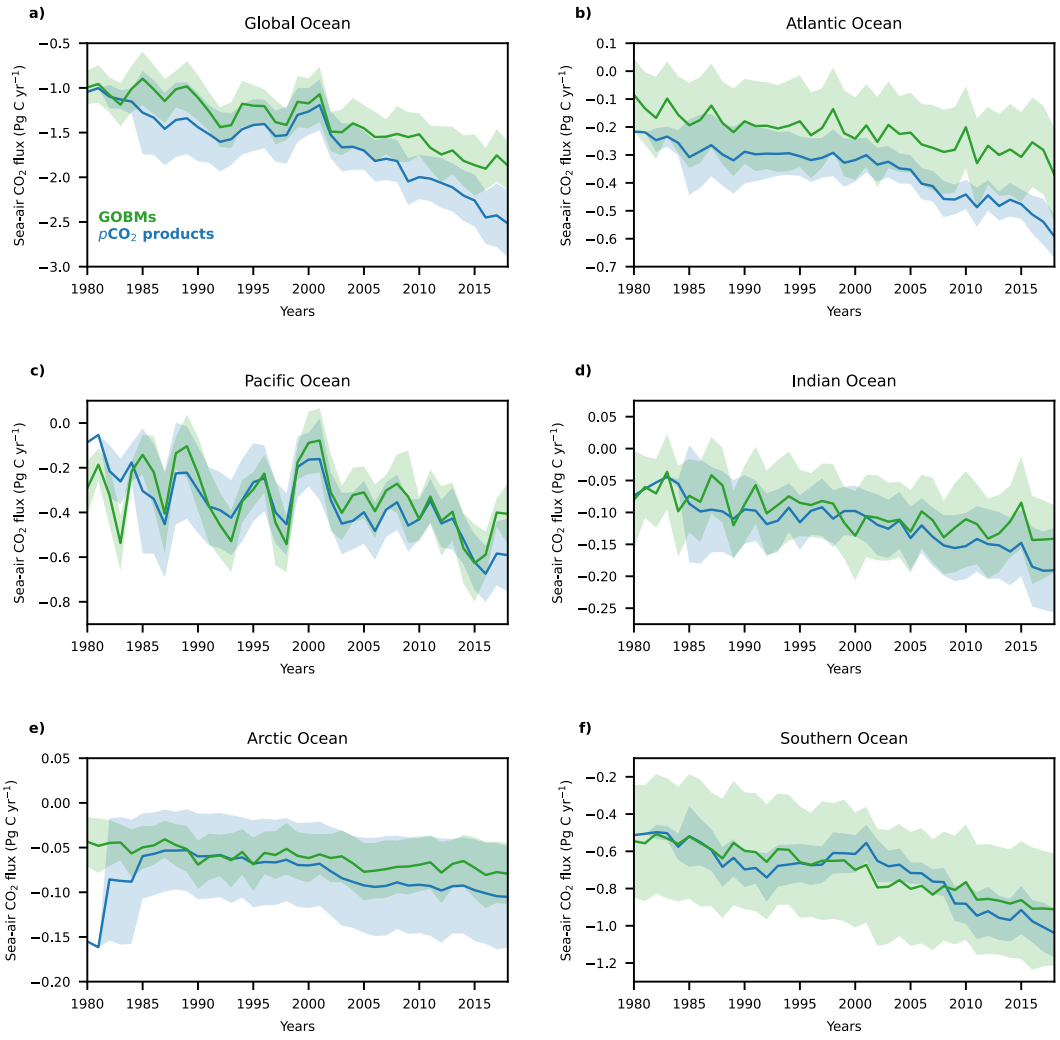


Figure 1. Time series of global and regional sea-air CO_2 fluxes from 1980 to 2018 based on GOBMs and pCO_2 products. *The average sea-air CO_2 flux from the GOBMs adjusted for the riverine-burial induced sea-air CO_2 flux (green) and pCO_2 products estimates (blue) for the* ***a)*** *global ocean , 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.*

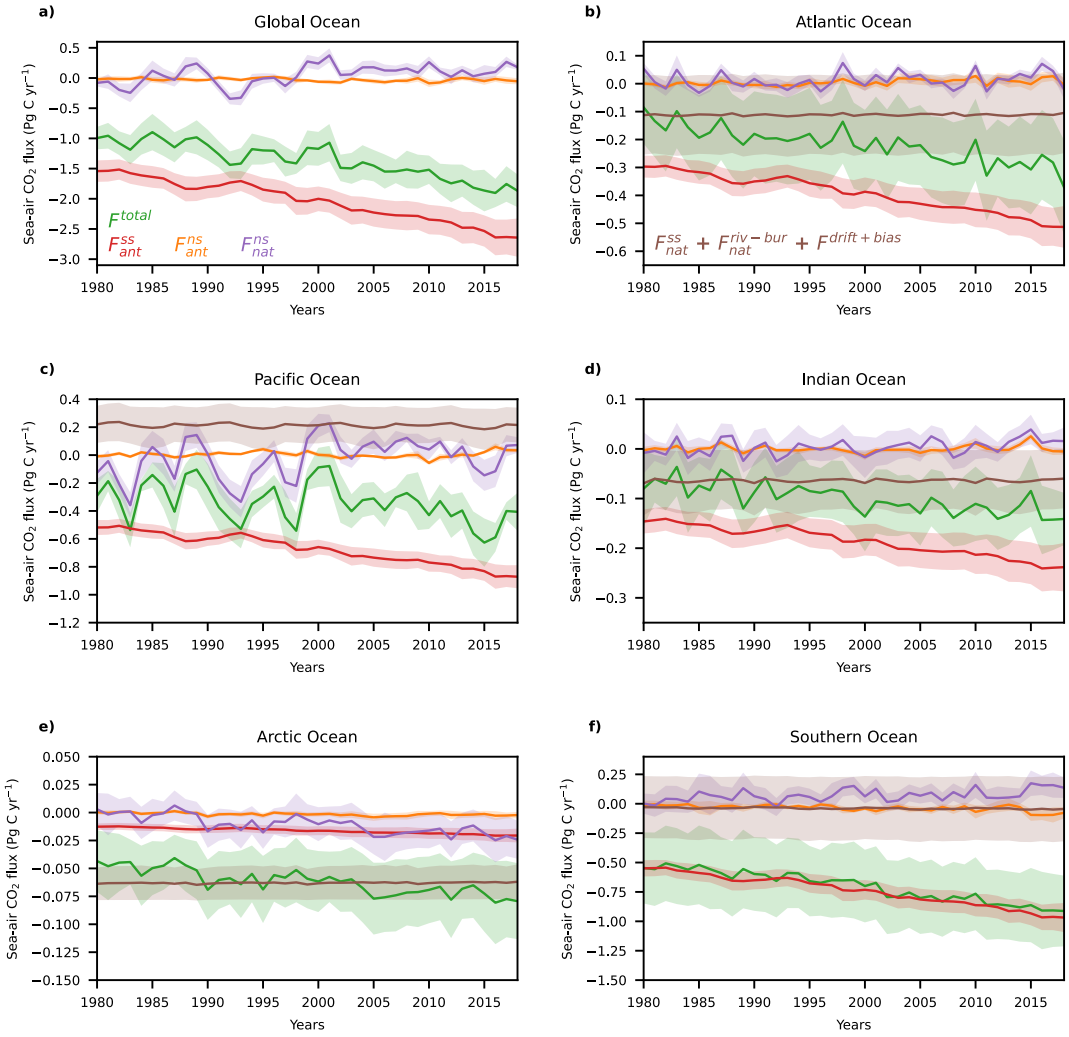


Figure 2. Time series of sea-air CO_2 flux components globally and regionally from 1980 to 2018 based on GOBMs. *The total sea-air CO_2 flux (F^{total}) integrated over each basin adjusted for the riverine-burial induced sea-air CO_2 flux (green) and the individual flux components from the GOBMs (F_{ant}^{ss} in red, F_{ant}^{ns} in orange, F_{nat}^{ns} in purple, and the sum of F_{nat}^{ss}, $F_{nat}^{riv\text{-}bur}$ and $F^{drift+bias}$ in brown) are shown for **a)** the global ocean and regionally for **b)** the Atlantic Ocean, **c)** the Pacific Ocean, **d)** the Indian Ocean, **e)** the Arctic Ocean, and **f)** the Southern Ocean. The shading indicates the respective standard deviation across all GOBMs. The uncertainty of F^{total} does not include the uncertainty of the riverine adjustment.*

501 ### 3.1 Sea-air CO_2 fluxes in the steady state control simulation

502 #### 3.1.1 Carbon fluxes from rivers and into sediments

503 The input of riverine carbon F_{nat}^{riv} and the sedimentation of carbon F_{nat}^{bur} is treated in
504 various ways across the ensemble of GOBMs and varies from 0.00 Pg C yr^{-1} to 0.61 PgC yr^{-1} and
505 from 0.00 Pg C yr^{-1} to 0.74 Pg C yr^{-1}, respectively (Table 2). The difference between F_{nat}^{riv} and
506 F_{nat}^{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
507 over the 8 GOBMs that provide all four simulations.

508

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

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

514

515 To estimate $F_{nat}^{riv\text{-}bur}$, we use the first-order assumption that $F_{nat}^{riv\text{-}bur} = F_{nat}^{riv} - F_{nat}^{bur}$ for all
516 GOBMs except NorESM-OC1.2 (Table 2). This assumption ignores the potential influence of
517 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 $F_{nat}^{riv\text{-}bur}$ is almost zero (Table 2). With the adjusted $F_{nat}^{riv\text{-}bur}$ for NorESM-OC1.2, the multi-model mean $F_{nat}^{riv\text{-}bur}$ is -0.03±0.15 Pg C yr^{-1}. In comparison, the model spread associated with F^{total} is 0.24 Pg C yr^{-1}. Although $F_{nat}^{riv\text{-}bur}$ does not directly affect the global estimation of F^{total}, it may substantially affect the regional estimates of F^{total} and F_{ss}^{nat}.

3.1.2 Bias and drift in the sea-air CO_2 flux due to incomplete spin-up

Across the ensemble of GOBMs, the approximated global $F^{drift+bias}$, quantified as the difference of F^{SimB} and $F_{nat}^{riv\text{-}bur}$ (equation 3, Table 2), varies from -0.35 to 0.17 Pg C yr^{-1}, with an ensemble mean of -0.06±0.18 Pg C yr^{-1}. The model spread around $F^{drift+bias}$ is of similar order as the model spread associated with the global F^{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 ±0.002 Pg C yr^{-2} for all GOBMs (Hauck et al., 2020). Although our estimation of $F^{drift+bias}$ is uncertain due to several approximations in our methodology, it gives a first indication of the importance of the non-steady-state for the model spread. A sufficiently long spin-up in each model to reach steady state may thus narrow down inter-model differences of regional F_{nat}^{ss} and F^{total}.

3.1.3 Steady state natural sea-air CO_2 flux

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

When comparing F^{SimB} in the Southern Ocean to F^{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 $F^{drift+bias}$ (-0.06±0.18 Pg C yr^{-1}) and $F_{nat}^{riv\text{-}bur}$ (-0.03±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 $F_{\text{nat}}^{\text{riv-bur}}$ (-0.09±0.22 Pg C yr^{-1} for the GOBM-ensemble without MOM6-Princeton) from the Southern Ocean F^{SimB} for each GOBM separately. This adjustment leads to an estimate of Southern Ocean $F_{\text{nat}}^{\text{ss}}$ 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^{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^{SimB} does not seem to be significantly impacted by the sum of $F^{\text{drift+bias}}$ and $F_{\text{nat}}^{\text{riv-bur}}$ across the GOBM ensemble or these fluxes cancel each other out. In these basins, we assume F^{SimB} to be approximately equal to $F_{\text{nat}}^{\text{ss}}$.

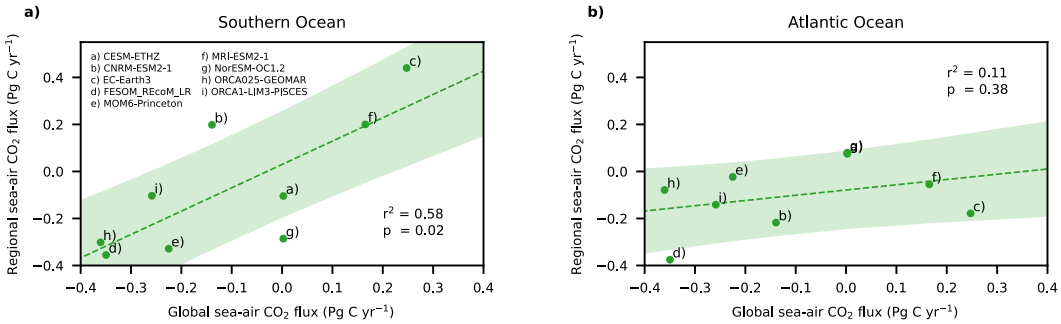


Figure 3. Relationship between global and regional sea-air CO_2 fluxes of simulation B for 9 GOBMs. *The relationship between sea-air CO_2 fluxes (averaged for 1980-2018, negative: into the ocean) of the global ocean and* ***a)*** *Southern Ocean and* ***b)*** *Atlantic Ocean is shown. Represented is 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.*

Our estimates of $F_{\text{nat}}^{\text{ss}}$ can be compared to inverse estimates of $F_{\text{nat}}^{\text{ss}}$ (Mikaloff Fletcher et al., 2007) (see also Section 2.3). These inverse estimates of $F_{\text{nat}}^{\text{ss}}$ 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 $F_{\text{nat}}^{\text{ss}}$ 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 $F^{drift+bias}$ and $F_{nat}^{riv-bur}$, the difference between the simulated and inverse-based estimate of F_{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_{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_{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_{nat}^{ns} carbon outgassing (Figure 4a, Figure 2). The average trend towards stronger outgassing of F_{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_{nat}^{ns} flux exist in the tropical Southern Ocean, the sea ice edge in the North Atlantic and Arctic Ocean, and the eastern coastal upwelling systems (Figure 4b).

The globally simulated inter-annual and decadal variability in F_{nat}^{ns} of 0.16±0.03 Pg C 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_{nat}^{ns} occurs in the tropical Pacific Ocean and the high-latitude oceans (Figure 4c). Though the pattern of variability is similar across the GOBMs, relatively large inter-model differences are found in the Southern Ocean, north-western Pacific Ocean, the North Atlantic subpolar gyre, and the Peruvian upwelling system (Figure 4d). The inter-annual and decadal variability in F_{nat}^{ns} is the dominant contributor to the inter-annual and decadal variability of F^{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_{ant}^{ns}) and regionally 2 to 6 times larger (Figure 2). The simulated temporal variability of F^{total} in the Pacific Ocean is driven by F_{nat}^{ns} (Figure 2c) and resembles the variability of F^{total} in the pCO_2 products (Figure 1). This good agreement indicates that the GOBMs represent the dominant source of Pacific sea-air CO_2 flux variability, El-Niño and La-Niña (Feely et al., 1999), well.

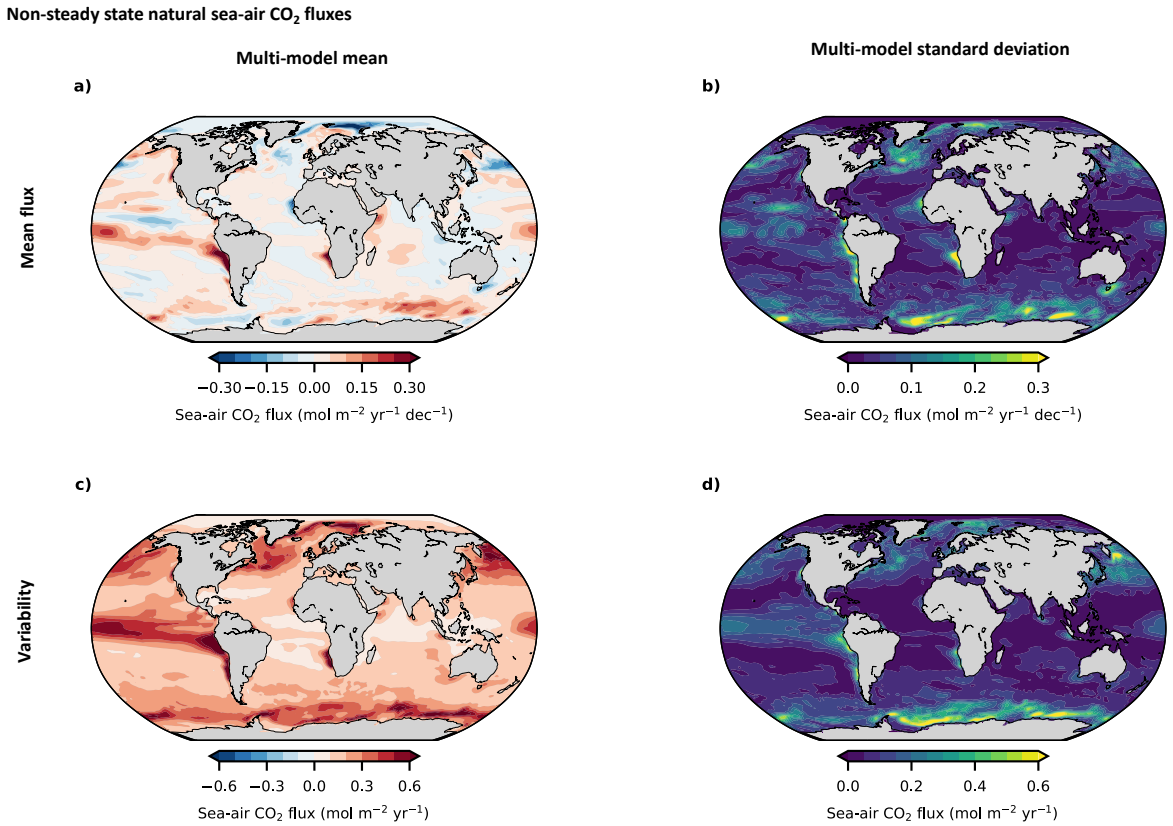


Figure 4. Non-steady state natural sea-air CO_2 fluxes for 8 GOBMs. *Maps of the* ***a)*** *the multi-model trend and* ***b)*** *the multi-model standard deviation of the trend of the natural non-steady state sea-air CO_2 flux from 1980 to 2018, as well as maps of* ***c)*** *the multi-model mean and* ***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 ΔC^*-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 ΔC^*-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°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_{ant} is 83±15 Pg C in 1994 , 22% (23 Pg C) lower than the ΔC^*-estimate, and 102±12 Pg C in 2002, 30% (44 Pg C) lower than the TTD-estimate. Most prominent differences are in the North Atlantic and Southern Ocean (Figure 5). These differences may be caused by the starting dates of the GOBM simulations that vary 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 ΔC^*-methodology might lead to an overestimation of C_{ant} in the upper water column and a negative bias in deeper waters (Matsumoto & Gruber, 2005).

As for the C_{ant}-storage since 1800, the spatial pattern of the simulated interior ocean C_{ant}-storage changes from 1994 to 2007 of the GOBMs resembles that of the eMLR(C^*)-estimate (Figure 5, Figure S4). 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}^{ss} by 20-30% as suggested by the Global Carbon Budget (Friedlingstein et al., 2022). Moreover, across the GOBM ensemble there exists a strong relationship between the simulated C_{ant} storage in 1994 since the beginning of the industrialization and the simulated change in C_{ant} storage from 1994 to 2007 across the model ensemble (Figure S5) 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}^{ss} by the GOBM ensemble.

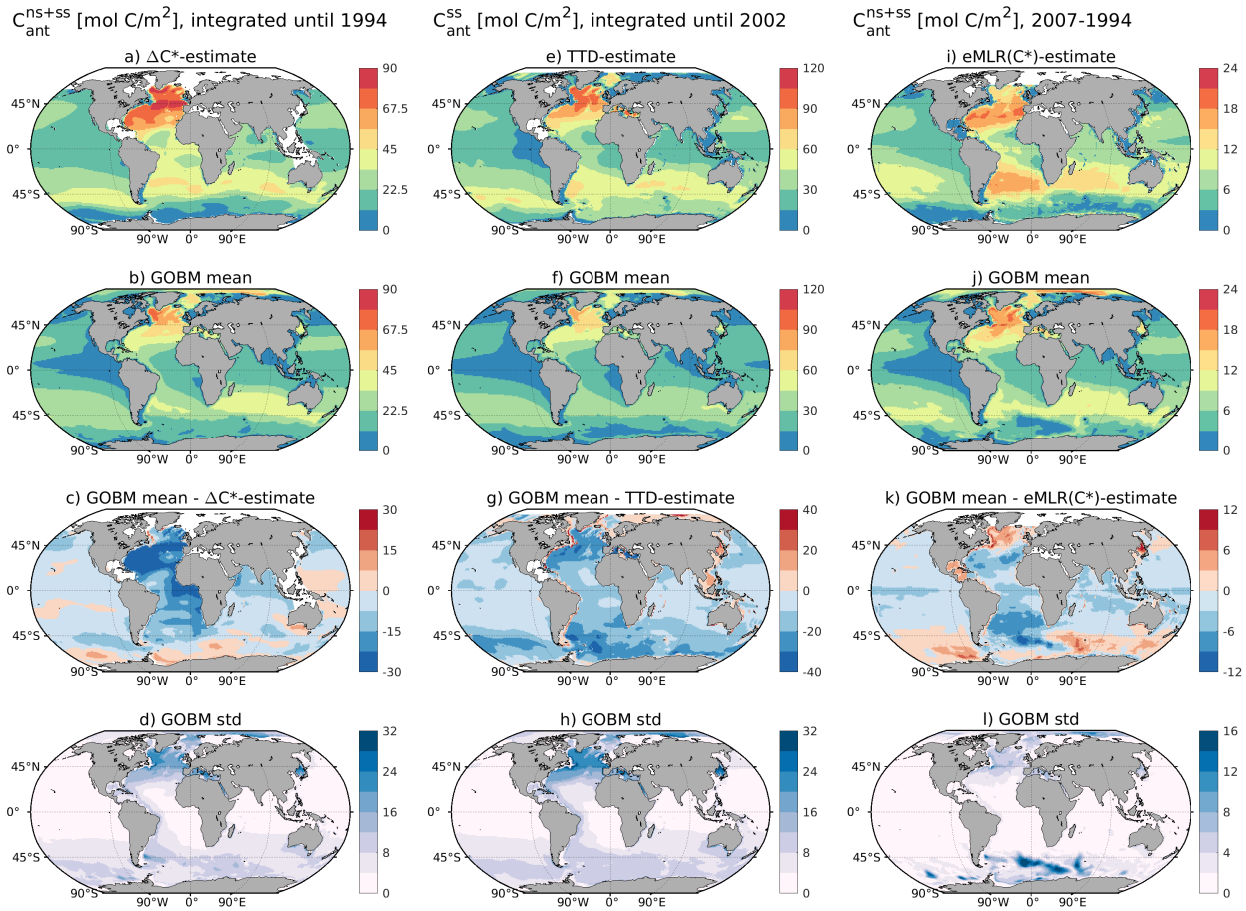


Figure 5: Column inventories of historic and contemporary anthropogenic carbon storage changes, integrated from surface to 3000m depth. *Visualised are* ***a) e) i)*** *observation-based estimates and related model-estimates based on 8 GOBMS, shown as* ***b) f) j)*** *model mean,* ***c) g) k)*** *difference 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}^{ns+ss} from the ΔC^*-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}^{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.

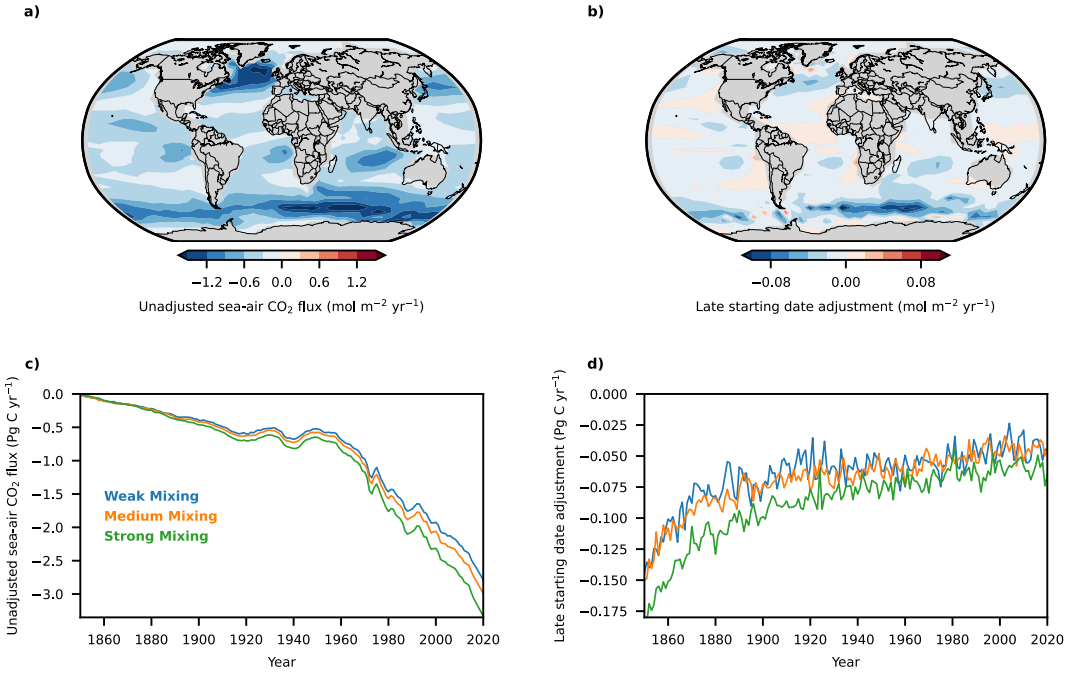


Figure 6: Difference in anthropogenic sea-air CO_2 fluxes due to different starting dates in Bern3D-LPX. *Maps of* ***a)*** *the anthropogenic sea-air CO_2 flux (steady-state and non-steady state) averaged from 1980 to 2018 and averaged over 3 Bern3D-LPX simulations with varying ocean mixing that start 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.*

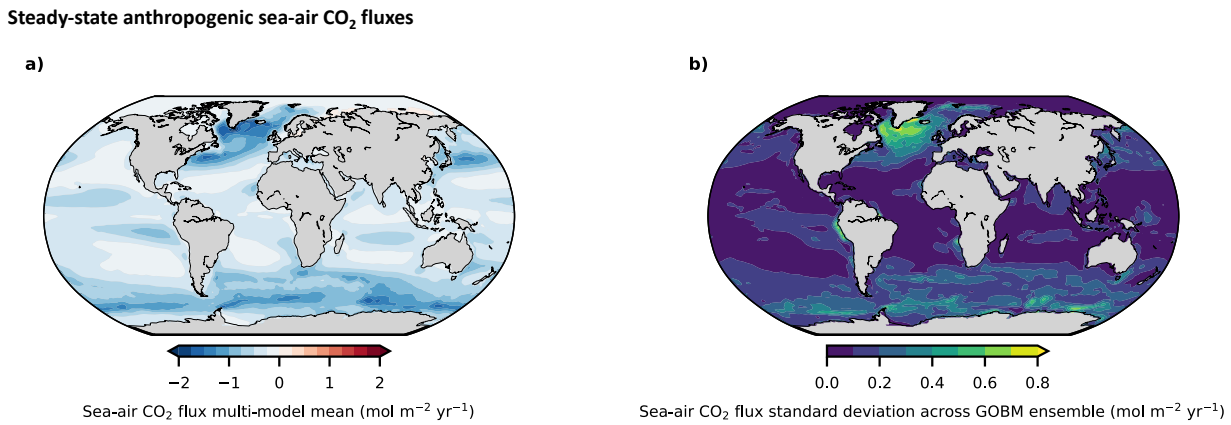
In comparison, the two offline approaches by Bronselaer et al. (2017) estimate an underestimation of the ocean carbon sink of 28.7±4.6 Pg C for the period from 1765 to 1995 when starting simulations in 1850 instead of 1765. More than 50% of this underestimation (~17 Pg C) is estimated to occur after 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

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

733



734

735 ***Figure 7: Simulated mean and intermodel spread of the steady-state anthropogenic CO_2 flux.***
736 *Maps of* ***a)*** *the multi-model mean and* ***b)*** *multi-model standard deviation of the steady state*
737 *anthropogenic sea-air CO_2 flux averaged from 1980 to 2018 for 8 GOBMs.*

738

739 3.3.3.1 Role of ocean circulation on steady state anthropogenic sea-air CO_2-fluxes in the
740 Atlantic and the Southern Ocean

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

749 Correlations between F_{ant}^{ss} and (i) the AMOC at 26.5°N or (ii) the storage of C_{ant} between
750 1000 m and 3000 m in the high latitude North Atlantic also occur across this ensemble of GOBMs
751 and can be used to identify emergent constraints (Figure 8a,b). In combination with the respective
752 observation-based estimates, the average annual Atlantic F_{ant}^{ss} from 1980 to 2018 can be
753 constrained from -0.39 ± 0.05 Pg C yr^{-1} to -0.43 ± 0.06 Pg C yr^{-1} when using the C_{ant}^{ss} storage and
754 to -0.42 ± 0.05 Pg C yr^{-1} when using the AMOC. The constraints identify a common bias in the
755 GOBMs towards too small AMOC strengths (mean underestimation of 18%) and C_{ant}^{ss} storage
756 below 1000m (mean underestimation of 22%), and hence Atlantic F_{ant}^{ss} (mean underestimation of

8-10%, depending on the used constraint). Nevertheless, the uncertainties around the Atlantic F_{ant}^{ss} estimate cannot be reduced due to the relatively large uncertainty of the observation-based estimate in case of the C_{ant}^{ss} storage as well as the relatively weak but significant correlation between the AMOC and the Atlantic F_{ant}^{ss} ($r^2 = 0.54$, $p = 0.04$). This weak correlation may partly be driven by the varying starting dates as GOBMs with a later or earlier starting date tend to have smaller or higher F_{ant}^{ss} than expected from the fit, respectively (Figure 8b). The correlation of the C_{ant}^{ss} storage and F_{ant}^{ss} is stronger ($r^2 = 0.84$, $p = 0.001$) because both variables are more directly related to each other and coherently affected by the late starting date. The relationships between Atlantic F_{ant}^{ss} and (i) AMOC and (ii) C_{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_{ant}^{ss} storage).

In the Southern Ocean, the magnitude of F_{ant}^{ss} also depends sensitively on the overturning circulation (Caldeira & Duffy, 2000; Mignone et al., 2006; Sarmiento et al., 1992), consisting here of the upwelling of circumpolar deep water close to the polar front, which is mainly transported northward, transferred to mode and intermediate waters, and eventually subducted at the subtropical front below the light subtropical surface waters into the ocean interior (Marshall & Speer, 2012; Talley, 2013). Across two ensembles of ESMs, it could be demonstrated that the volume of ventilated mode and intermediate waters in the Southern Ocean is highly correlated with the sea surface density between the polar front and the subtropical front, 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_{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_{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_{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_{ant}^{ss} can partly be explained by different starting dates as GOBMs with a late or early starting date have a smaller or larger absolute F_{ant}^{ss} than expected from the linear fit between the mean surface salinity in the subtropical-polar frontal zone, respectively (Figure 8c). A common starting date 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_{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_{ant}^{ss} from 1980 to 2018 in Atlantic and Southern Ocean only slightly, the influence of circulation biases on F_{ant}^{ss} increases in magnitude with increasing atmospheric F_{ant}^{ss}. Therefore, the difference between constrained and unconstrained F_{ant}^{ss} increases over time (Figure S6) and a GOBM ensemble with circulation biases will have smaller trends in F_{ant}^{ss} and deviate from the true F_{ant}^{ss} with time.

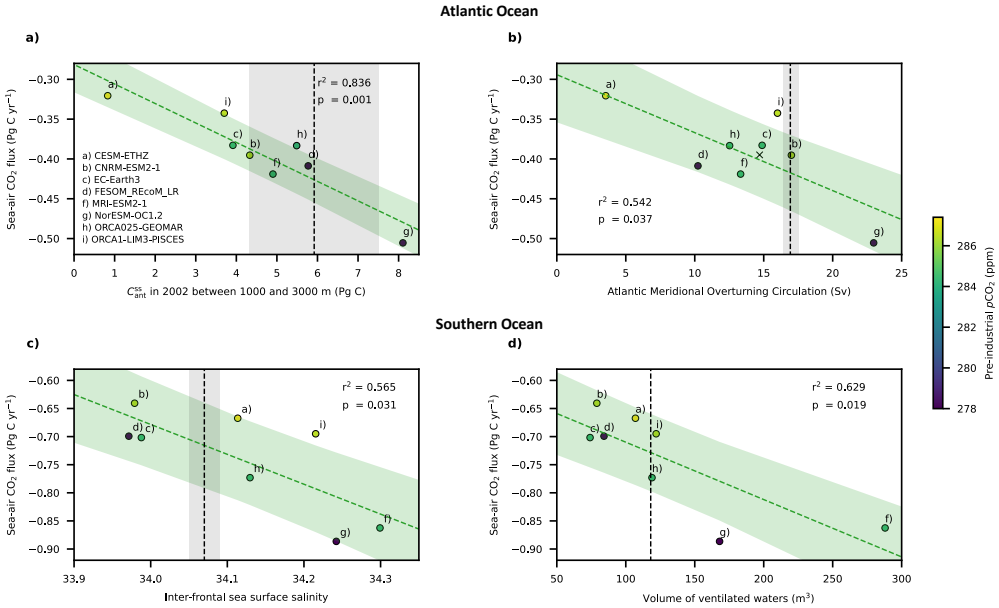


Figure 8. Constrained steady-state anthropogenic carbon uptake in the Atlantic and Southern Ocean. *Steady-state anthropogenic carbon uptake averaged from 1980 to 2018 of* ***a)b)*** *the Atlantic and* ***c)d)*** *the Southern Ocean, plotted against* ***a)*** *the Atlantic steady-state anthropogenic carbon storage between 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

(Figure 9a). The GOBMs reproduce this pattern on average (Figure 9b) but show larger chemical uptake capacities in the tropical and subtropical oceans, and smaller chemical uptake capacities in the subpolar gyres, most of the Southern Ocean, the Labrador Sea, and the Arctic Ocean (Figure 9c). The inter-model variability is small in most places apart from sea ice regions in the Arctic Ocean and in eastern upwelling systems west of South America and Africa (Figure 9d), suggesting common biases in the chemical uptake capacities across the GOBM ensemble.

Globally, the chemical uptake capacity of the eight GOBMs is similar to that of the pCO_2/alkalinity products and of GLODAPv2 (Figure 9e). This capacity is directly linked to the surface alkalinity (Figure 9h) as GOBMs with a high buffer capacity have also high surface ocean CO_3^{2-} concentrations (Figure 9f), a high difference in surface 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_{ant}^{ss} (Figure 9e), although studies with ESMs found such a relationship (Terhaar et al., 2022). Possible reasons for no emerging relationship between F_{ant}^{ss} and the chemical uptake capacity, CO_3^{2-}, or the alkalinity across the GOBM ensemble are differences in F_{ant}^{ss} due to different starting dates of the simulations (Section 3.3.2) and ongoing $F^{drift+bias}$. If a GOBM has a large negative or positive $F^{drift+bias}$, its upwelling waters have too low or high DIC, too high or low CO_3^{2-}, and hence a chemical uptake capacity that is too high or low, respectively. With time, the additional surface ocean DIC from $F^{drift+bias}$ reduces the chemical uptake capacity so that it is effectively smaller than the one expected from the theoretical chemical uptake capacity. Thus, $F^{drift+bias}$ adds considerable noise so that a potential relationship between the chemical uptake capacity and F_{ant}^{ss} may not be identifiable. When considering only the four GOBMs with a longer spin-up than 1000 years, a relationship indeed emerges (Figure 9e-h).

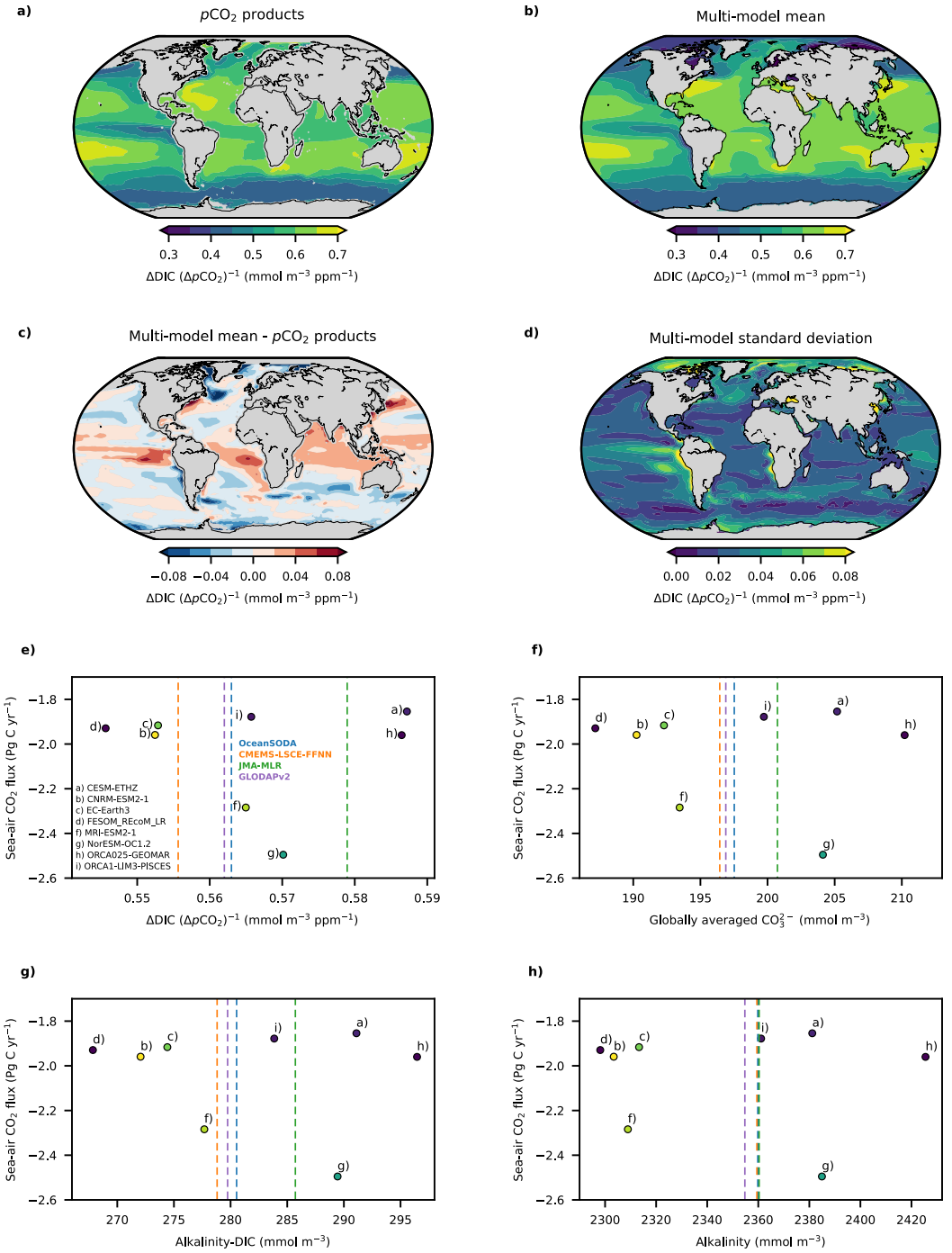


Figure 9: Surface ocean chemical uptake capacity and its relationship to the steady-state anthropogenic sea-air CO_2 flux. *Maps of the increase in DIC per increase in pCO_2 averaged from 1986 to 2018 based on* ***a)*** *3 pCO_2/alkalinity products (average of OceanSODA-ETHZ, CMEMS-LSCE-FFNN, and JMA-MLR)) 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)***

869 *increase in DIC per increase in pCO_2,* ***f)*** *surface ocean CO_3^{2-} concentration,* ***g)*** *difference between*
870 *surface ocean alkalinity and DIC, and* ***h)*** *the global surface ocean alkalinity. The colors of each*
871 *dot that represents a GOBM indicate the number of simulated years before the start of the analyzed*
872 *period in 1980, and the dashed lines indicate each pCO_2/alkalinity product and GLODAPv2 for*
873 *the variables on the respective x-axis.*

874

875 3.4 Non-steady state anthropogenic sea-air CO_2 flux

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

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

905

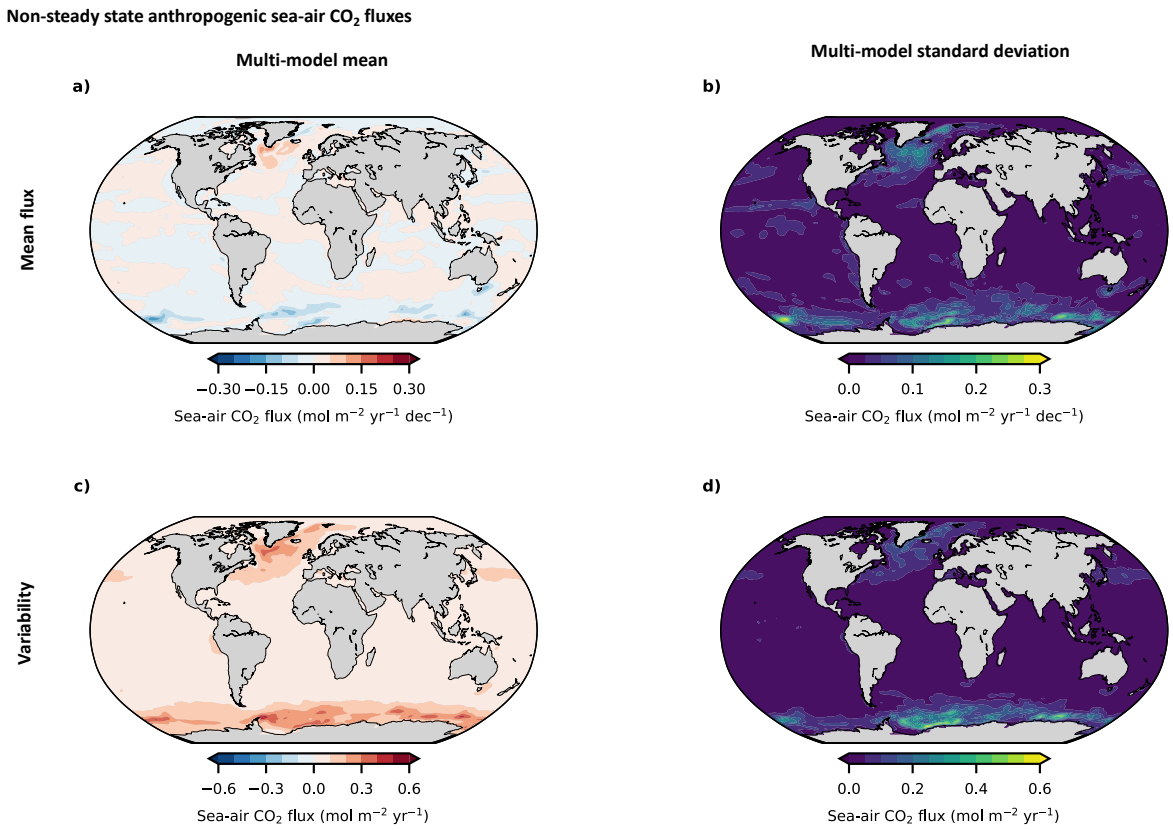


Figure 10. Non-steady state anthropogenic sea-air CO_2 fluxes for 8 GOBMs. *Maps of the* ***a)*** *the multi-model mean and* ***b)*** *the multi-model standard deviation of the linear trend in anthropogenic non-steady state sea-air CO_2 flux without the inter-annual variability (calculated by fitting 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^{total} (based on equation (9)) as $F^{\text{drift+bias}}$ is removed when subtracting F^{SimB} from F^{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^{total}. Regionally, $F^{\text{drift+bias}}$ directly affects F^{total} because subtracting F^{SimB} from F^{SimA} removes not only $F^{\text{drift+bias}}$ but also $F_{\text{nat}}{}^{\text{ss}}$, which is regionally not zero. To regionally estimate F^{total} from a GOBM, one could hence rely either on F^{SimA} - F^{SimB} and add an independent estimate of $F_{\text{nat}}{}^{\text{ss}}$ (e.g., the inverse model estimate from Mikaloff Fletcher et al. (2007)), which comes with its own uncertainties, or rely on F^{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^{total} in that ocean region. When assuming that $F^{drift+bias}$ is almost entirely located in the Southern Ocean (Section 3.3.1), $F^{drift+bias}$ 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 $F^{drift+bias}$ (Figure 11) suggests that a short spin-up is often insufficient to reduce $F^{drift+bias}$ (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^{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.

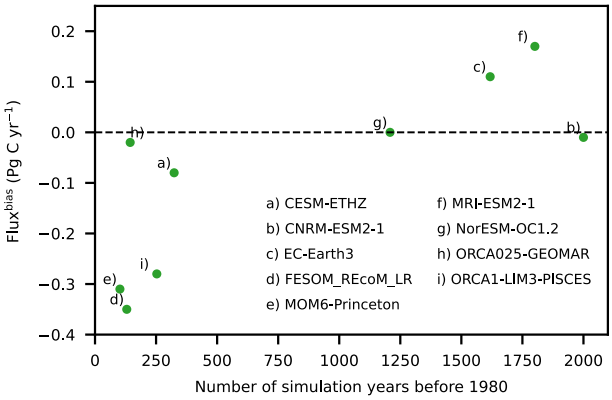


Figure 11. Estimated global sea-air CO_2 bias fluxes related to the models not being in steady-state for 9 GOBMs against the length of their spin-up. *The length of the spin-up is defined as the number of simulated years at that resolution before the start of the analyzed period in 1980, while the bias-flux (F^{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 $F_{nat}^{riv\text{-}bur}$ comparable to the observation-based $F_{obs}^{riv\text{-}bur}$ 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^{total} are however unaffected by $F_{nat}^{riv\text{-}bur}$ if equation (9) is used as the simulated $F_{nat}^{riv\text{-}bur}$ is removed when subtracting F^{SimB} from F^{SimA} and replaced by the observation-based estimate of riverine and sediment fluxes. 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^{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 $F_{obs}^{riv\text{-}bur}$ of -0.65±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^{total} before accounting for $F_{obs}^{riv\text{-}bur}$ (±0.24 Pg C yr^{-1}).

Regionally, the uncertainties of $F_{obs}^{riv\text{-}bur}$ are even larger than globally. Across RECCAP2 chapters, the local distribution of $F_{obs}^{riv\text{-}bur}$ 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 $F_{obs}^{riv\text{-}bur}$ in the Atlantic Ocean and a larger $F_{obs}^{riv\text{-}bur}$ in the Southern Ocean. One reason for this difference might be that Lacroix et al. (2021) quantify $F_{obs}^{riv\text{-}bur}$ 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 $F_{obs}^{riv\text{-}bur}$ 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 $F_{obs}^{riv\text{-}bur}$ 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^{total} in the Atlantic Ocean would double. Hence, more refined estimates of the lability of organic matter and its effect on $F_{obs}^{riv\text{-}bur}$ are crucial to better constrain the 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_{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_{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_{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_{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_{ant}^{ss} (Section 3.4.5.1), i.e., increasing F_{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_{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^{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 $F_{obs}^{riv\text{-}bur}$ estimate). The GOBM underestimation of 0.30±0.38 Pg C yr^{-1} (~18% of the F^{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_{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_{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^{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^{total} would not appear as a divergence of both estimates since 2000 but as a change from an underestimation of F^{total} by the

1085 pCO_2 products to an overestimation. Nevertheless, the growth rates of F^{total} are different between
1086 GOBMs and pCO_2 products and uncertainties remain of how the ocean sink evolves.

1087 Regionally, different trends in F^{total} between GOBMs and pCO_2 products seem to be driven
1088 by a mismatch in the temporal evolution of the Southern Ocean carbon sink, and an increasing gap
1089 between both estimates in the Atlantic (Figure 1). In the Southern Ocean, the pCO_2 product
1090 estimate of the Southern Ocean carbon sink suggested that the variability before 2000 is mainly
1091 due to decadal variations (Gruber, Landschützer, et al., 2019; Keppler & Landschützer, 2019;
1092 Landschützer et al., 2015; McKinley et al., 2017). Since 2000, the estimate of the pCO_2 products
1093 of the Southern Ocean carbon flux has been moving toward more uptake. While this ongoing
1094 increase in uptake based on the pCO_2 products of the Southern Ocean may just be a longer
1095 variability cycle, it could also indicate a disagreement on the trend of the ocean carbon sink
1096 between pCO_2-based and GOBM-based estimates for unknown reasons. Moreover, it remains an
1097 open question if differences between both estimates are due to the erroneous models or the
1098 extrapolation of sparse observations with temporal aliasing.

1099 The increasing gap in the Atlantic after 2000, however, appears to result from a smaller
1100 F^{total} trend in GOBMs than in pCO_2 products. This smaller-than-observed trend in GOBMs can
1101 partly be explained by the negatively biased chemical uptake capacity of the GOBMs (Section
1102 3.3.3.2). Related to this, Lebehot et al. (2019) showed for a suite of ESMs that the North Atlantic
1103 surface ocean fugacity of CO_2 increased at a significantly faster rate than observed and related this
1104 to substantial biases in alkalinity and its impact on the buffer capacity. The GOBMs also show a
1105 biased-small AMOC, whose influence on F_{ant}^{ss} increases with increasing atmospheric CO_2
1106 (Section 3.3.3.1; Figure S6). Perez et al. (to be submitted) show that the disagreement in Atlantic
1107 F^{total} trends between GOBMs and pCO_2 products is especially large in the subpolar North Atlantic.
1108 This relates well to our finding about AMOC-biases as the influence of AMOC-biases on F_{ant}^{ss} is
1109 potentially highest in the subpolar gyre where subsurface waters low in C_{ant} outcrop. Furthermore,
1110 a study with ESMs has shown that AMOC-biases are strongly correlated to SST-biases in the
1111 North Atlantic (Wang et al., 2014). While we did not analyze SST biases in the North Atlantic,
1112 Rodgers et al. (in review) found that the seasonal cycle of pCO_2 in the subpolar Atlantic is
1113 thermally driven in the GOBMs while that of the pCO_2-products is non-thermally driven. This
1114 might lead to the F^{total} of the GOBMs being more sensitive to warming (Goris et al., 2018), which
1115 may contribute to the increasing gap between GOBMs and pCO_2-products with time. However,
1116 the magnitude of these contributions is unclear and remains to be identified.

1117

1118 ### 4.7 Inter-annual and decadal variability of the sea-air CO_2 flux

1119 The here-used GOBM simulations suggest that, for the time-period 1980-2018, the largest
1120 share of the inter-annual and decadal variability of F^{total} results from F_{nat}^{ns}, i.e., the sea-air flux of
1121 natural carbon due to climate variability and climate change. Globally, F_{nat}^{ns} is also an important
1122 flux component as it allows comparing the estimated ocean carbon sink from surface ocean pCO_2
1123 products, which quantify F_{ant}^{ss}, F_{ant}^{ns}, F_{nat}^{ns}, and $F_{obs}^{riv\text{-}bur}$ (Friedlingstein et al., 2022) to
1124 observation-based estimates of the interior ocean change of C_{ant} (Gruber, Clement, et al., 2019),
1125 which quantifies only changes in F_{ant}^{ss}, and F_{ant}^{ns}.

Previous approximations estimated the global F_{nat}^{ns} from 1994 to 2007 to be 5±3 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_{nat}^{ns} of 1.6±0.8 Pg C over the same period, which is similar to both previous model-based estimates, although the ESM-based estimate accounts only for the effect of climate change and externally forced variability (volcanoes, variability in atmospheric CO_2) and not for the unforced variability of the climate system (e.g. winds, atmospheric temperature etc).

Regionally, the variability of the sea-air CO_2 flux is similar between GOBMs and pCO_2 products in the Pacific Ocean, where most of the inter-annual variability is located, and differs in the Southern Ocean, where pCO_2 products suggest a strong decadal variability before 2000 and a 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 pCO_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_{nat}^{ns}, the largest F_{ant}^{ns} are simulated in the subpolar North Atlantic with yet unidentified drivers and in the Southern Ocean where sea ice retreats with global warming and westerly winds strengthen and shift southwards (Purich et al., 2016). The strengthening of F_{ant}^{ns} in the Southern Ocean could be explained by additional free ocean surface due to climate change, which can thus take up more C_{ant} or by more upwelling of old water with low C_{ant} content (Le Quéré et al., 2007), which can also take up more C_{ant}. Both processes would lead to partial compensation by F_{nat}^{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_{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 pCO_2 from SOCAT (Bakker et al., 2016), which suggests an underestimation of the ocean carbon uptake by GOBMs. This is consistent with our assessment that suggests an underestimation of the simulated ocean carbon 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 pCO_2 products flux estimates regionally, thereby repeating an analysis from the RECCAP1 project by Séférian et al. (2014). By selecting the GOBMs that perform best, they suggest that the simulated global

ocean carbon sink is smaller than previously estimated, opposite to what this study and Hauck et al. (2020) suggest. Several assumptions are made by Fay & McKinley (2021), such as the application of the local riverine adjustment by Lacroix et al. (2021), not accounting for each models' simulated regional $F_{nat}^{riv\text{-}bur}$, and that an area-weighted repartitioning $F^{drift+bias}$ over the entire ocean surface is valid. However, the local riverine adjustments come with large uncertainties (Section 4.2) and our analysis suggests that $F^{drift+bias}$ and $F_{nat}^{riv\text{-}bur}$ 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 ΔC^* 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±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^{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_{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

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

Aumont, O., Ethé, C., Tagliabue, A., Bopp, L., & Gehlen, M. (2015). PISCES-v2: an ocean

biogeochemical model for carbon and ecosystem studies. *Geoscientific Model Development*, *8*(8), 2465–2513. https://doi.org/10.5194/gmd-8-2465-2015

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

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

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

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

Broecker, W. S., TTakahashi, T., Simpson, H. J., & Peng T.-H. (1979). Fate of Fossil Fuel Carbon Dioxide and the Global Carbon Budget. *Science*, *206*(4417), 409–418. https://doi.org/10.1126/science.206.4417.409

Bronselaer, B., Winton, M., Russell, J., Sabine, C. L., & Khatiwala, S. (2017). Agreement of CMIP5 Simulated and Observed Ocean Anthropogenic CO_2 Uptake. *Geophysical Research Letters*, *44*(24), 12, 212–298, 305. https://doi.org/https://doi.org/10.1002/2017GL074435

Caldeira, K., & Duffy, P. B. (2000). The Role of the Southern Ocean in Uptake and Storage of Anthropogenic Carbon Dioxide. *Science*, *287*(5453), 620–622. https://doi.org/10.1126/science.287.5453.620

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

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

Clement, D., & Gruber, N. (2018). The eMLR(C^*) Method to Determine Decadal Changes in the Global Ocean Storage of Anthropogenic CO2. *Global Biogeochemical Cycles*, *32*(4), 654–679. https://doi.org/https://doi.org/10.1002/2017GB005819

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

Davila, X., Gebbie, G., Brakstad, A., Lauvset, S. K., McDonagh, E. L., Schwinger, J., & Olsen, A. (2022). How Is the Ocean Anthropogenic Carbon Reservoir Filled? *Global Biogeochemical Cycles*, *36*(5), e2021GB007055. https://doi.org/https://doi.org/10.1029/2021GB007055

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

DeVries, T. (2022). Atmospheric CO_2 and Sea Surface Temperature Variability Cannot Explain Recent Decadal Variability of the Ocean CO_2 Sink. *Geophysical Research Letters*, *49*(7), e2021GL096018. https://doi.org/https://doi.org/10.1029/2021GL096018

DeVries, T., Yamamoto, K., Wanninkhof, R., Gruber, N., Hauck, J., Müller, J. D., et al. (in review). Magnitude, trends, and variability of the global ocean carbon sink from 1985-2018. *Global Biogeochemical Cycles*.

Dickson, A. G., Sabine, C. L., & Christian, J. R. (2007). *Guide to best practices for ocean CO2 measurements.* North Pacific Marine Science Organization.

Döscher, R., Acosta, M., Alessandri, A., Anthoni, P., Arsouze, T., Bergman, T., et al. (2022). The EC-Earth3 Earth system model for the Coupled Model IntercomparisonProject 6. *Geoscientific Model Development*, *15*(7), 2973–3020. https://doi.org/10.5194/gmd-15-2973-2022

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

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

Fay, A. R., & McKinley, G. A. (2021). Observed Regional Fluxes to Constrain Modeled Estimates of the Ocean Carbon Sink. *Geophysical Research Letters*, *48*(20), e2021GL095325. https://doi.org/https://doi.org/10.1029/2021GL095325

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

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

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

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

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

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

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

Fu, W., Moore, J. K., Primeau, F., Collier, N., Ogunro, O. O., Hoffman, F. M., & Randerson, J. T. (2022). Evaluation of Ocean Biogeochemistry and Carbon Cycling in CMIP Earth System Models With the International Ocean Model Benchmarking (IOMB) Software System. *Journal of Geophysical Research: Oceans*, *127*(10), e2022JC018965.

https://doi.org/https://doi.org/10.1029/2022JC018965
Gao, S., Schwinger, J., Tjiputra, J., Bethke, I., Hartmann, J., Mayorga, E., & Heinze, C. (2023). Riverine impact on future projections of marine primary production and carbon uptake. *Biogeosciences*, *20*(1), 93–119. https://doi.org/10.5194/bg-20-93-2023
Gattuso, J.-P., Magnan, A. K., Bopp, L., Cheung, W. W. L., Duarte, C. M., Hinkel, J., et al. (2018). Ocean Solutions to Address Climate Change and Its Effects on Marine Ecosystems.*Frontiers in Marine Science*. Retrieved from https://www.frontiersin.org/articles/10.3389/fmars.2018.00337
Gloege, L., McKinley, G. A., Landschützer, P., Fay, A. R., Frölicher, T. L., Fyfe, J. C., et al. (2021). Quantifying Errors in Observationally Based Estimates of Ocean Carbon Sink Variability. *Global Biogeochemical Cycles*, *35*(4), e2020GB006788. https://doi.org/https://doi.org/10.1029/2020GB006788
Gomez, F. A., Wanninkhof, R., Barbero, L., & Lee, S.-K. (2021). Increasing River Alkalinity Slows Ocean Acidification in the Northern Gulf of Mexico. *Geophysical Research Letters*, *48*(24), e2021GL096521. https://doi.org/https://doi.org/10.1029/2021GL096521
Goris, N., Tjiputra, J., Schwinger, J., & Heinze, C. (2015). Responses of carbon uptake and oceanic pCO_2 to climate change in the North Atlantic: A model study with the Bergen Earth System Model. *Global Biogeochemical Cycles*, *29*(10), 1567–1583. https://doi.org/https://doi.org/10.1002/2015GB005109
Goris, N., Tjiputra, J. F., Olsen, A., Schwinger, J., Lauvset, S. K., & Jeansson, E. (2018). Constraining Projection-Based Estimates of the Future North Atlantic Carbon Uptake. *Journal of Climate*, *31*(10), 3959–3978. https://doi.org/10.1175/JCLI-D-17-0564.1
Goris, N., Johannsen, K., & Tjiputra, J. (2023). The emergence of the Gulf Stream and interior western boundary as key regions to constrain the future North Atlantic carbon uptake. *Geoscientific Model Development*, *16*(8), 2095–2117. https://doi.org/10.5194/gmd-16-2095-2023
Gregor, L., & Gruber, N. (2021). OceanSODA-ETHZ: a global gridded data set of the surface ocean carbonate system for seasonal to decadal studies of ocean acidification. *Earth System Science Data*, *13*(2), 777–808. https://doi.org/10.5194/essd-13-777-2021
Gregor, L., Lebehot, A. D., Kok, S., & Scheel Monteiro, P. M. (2019). A comparative assessment of the uncertainties of global surface ocean CO_2 estimates using a machine-learning ensemble (CSIR-ML6 version2019a) – have we hit the wall? *Geoscientific Model Development*, *12*(12), 5113–5136. https://doi.org/10.5194/gmd-12-5113-2019
Griffies, S. M., Danabasoglu, G., Durack, P. J., Adcroft, A. J., Balaji, V., Böning, C. W., et al. (2016). OMIP contribution to CMIP6: experimental and diagnostic protocol for the physical component of the Ocean Model Intercomparison Project. *Geoscientific Model Development*, *9*(9), 3231–3296. https://doi.org/10.5194/gmd-9-3231-2016
Gruber, N., Sarmiento, J. L., & Stocker, T. F. (1996). An improved method for detecting anthropogenic CO_2 in the oceans. *Global Biogeochemical Cycles*, *10*(4), 809–837. https://doi.org/https://doi.org/10.1029/96GB01608
Gruber, N., Clement, D., Carter, B. R., Feely, R. A., van Heuven, S., Hoppema, M., et al. (2019). The oceanic sink for anthropogenic CO_2 from 1994 to 2007. *Science*, *363*(6432), 1193–1199. https://doi.org/10.1126/science.aau5153
Gruber, N., Landschützer, P., & Lovenduski, N. S. (2019, January 3). The variable southern ocean carbon sink. *Annual Review of Marine Science*. Annual Reviews. https://doi.org/10.1146/annurev-marine-121916-063407

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

Gürses, Ö., Oziel, L., Karakus, O., Sidorenko, D., Völker, C., Ye, Y., et al. (2023). Ocean biogeochemistry in the coupled ocean-sea ice-biogeochemistry model FESOM2.1-REcoM3. *Geoscientific Model Development Discussions*, *2023*, 1–73. https://doi.org/10.5194/gmd-2023-2

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

Hauck, J., Zeising, M., Le Quéré, C., Gruber, N., Bakker, D. C. E., Bopp, L., et al. (2020). Consistency and Challenges in the Ocean Carbon Sink Estimate for the Global Carbon Budget. *Frontiers in Marine Science*, *7*. https://doi.org/10.3389/fmars.2020.571720

Hauck, J., Gregor, L., Nissen, C., Patara, L., Hague, M., Mongwe, N. P., et al. (to be submitted). The Southern Ocean carbon cycle 1985-2018: mean, seasonal cycle, trends and storage. *Global Biogeochemical Cycles*.

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

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

IPCC (2021). Summary for Policymakers. In V. Masson-Delmotte, P. Zhai, A. Pirani, S. L. Connors, C. Péan, S. Berger, et al. (Eds.), *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press.

Joos, F., Plattner, G.-K., Stocker, T. F., Marchal, O., & Schmittner, A. (1999). Global Warming and Marine Carbon Cycle Feedbacks on Future Atmospheric CO_2. *Science*, *284*(5413), 464–467. https://doi.org/10.1126/science.284.5413.464

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

Keppler, L., & Landschützer, P. (2019). Regional Wind Variability Modulates the Southern Ocean Carbon Sink. *Scientific Reports*, *9*(1), 7384. https://doi.org/10.1038/s41598-019-43826-y

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

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

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

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

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

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

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

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

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

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

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

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

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

Louchard, D., Gruber, N., & Münnich, M. (2021). The Impact of the Amazon on the Biological Pump and the Air-Sea CO_2 Balance of the Western Tropical Atlantic. *Global Biogeochemical Cycles*, *35*(6), e2020GB006818. https://doi.org/https://doi.org/10.1029/2020GB006818

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

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

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

Marshall, J., & Speer, K. (2012). Closure of the meridional overturning circulation through

Southern Ocean upwelling. *Nature Geoscience*, *5*(3), 171–180. https://doi.org/10.1038/ngeo1391

Matear, R. J., Wong, C. S., & Xie, L. (2003). Can CFCs be used to determine anthropogenic CO2? *Global Biogeochemical Cycles*, *17*(1). https://doi.org/https://doi.org/10.1029/2001GB001415

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

Mauritsen, T., Bader, J., Becker, T., Behrens, J., Bittner, M., Brokopf, R., et al. (2019). Developments in the MPI-M Earth System Model version 1.2 (MPI-ESM1.2) and Its Response to Increasing CO_2. *Journal of Advances in Modeling Earth Systems*, *11*(4), 998–1038. https://doi.org/https://doi.org/10.1029/2018MS001400

McCarthy, G. D., Smeed, D. A., Johns, W. E., Frajka-Williams, E., Moat, B. I., Rayner, D., et al. (2015). Measuring the Atlantic Meridional Overturning Circulation at 26°N. *Progress in Oceanography*, *130*, 91–111. https://doi.org/https://doi.org/10.1016/j.pocean.2014.10.006

McKinley, G. A., Fay, A. R., Lovenduski, N. S., & Pilcher, D. J. (2017). Natural Variability and Anthropogenic Trends in the Ocean Carbon Sink. *Annual Review of Marine Science*, *9*(1), 125–150. https://doi.org/10.1146/annurev-marine-010816-060529

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

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

Middelburg, J. J., Soetaert, K., & Hagens, M. (2020). Ocean Alkalinity, Buffering and Biogeochemical Processes. *Reviews of Geophysics*, *58*(3), e2019RG000681. https://doi.org/https://doi.org/10.1029/2019RG000681

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

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

Mikaloff Fletcher, S. E., Gruber, N., Jacobson, A. R., Gloor, M., Doney, S. C., Dutkiewicz, S., et al. (2007). Inverse estimates of the oceanic sources and sinks of natural CO_2 and the implied oceanic carbon transport. *Global Biogeochemical Cycles*, *21*(1). https://doi.org/https://doi.org/10.1029/2006GB002751

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

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

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

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

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

Perez, F. F., Gehlen, M., Tjiputra, J. F., Olsen, A., Becker, M., Goris, N., et al. (to be submitted). An assessment of CO_2 storage and sea-air fluxes for the Atlantic Ocean and Mediterranean Sea between 1985 and 2018. *Global Biogeochemical Cycles*.

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

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

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

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

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

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

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

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

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

Revelle, R., & Suess, H. E. (1957). Carbon Dioxide Exchange Between Atmosphere and Ocean and the Question of an Increase of Atmospheric CO_2 during the Past Decades. *Tellus*, *9*(1), 18–27. https://doi.org/https://doi.org/10.1111/j.2153-3490.1957.tb01849.x

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

Rödenbeck, C., Keeling, R. F., Bakker, D. C. E., Metzl, N., Olsen, A., Sabine, C., & Heimann, M. (2013). Global surface-ocean pCO_2 and sea–air CO_2 flux variability from an observation-driven ocean mixed-layer scheme. *Ocean Science*, *9*(2), 193–216. https://doi.org/10.5194/os-9-193-2013

Rodgers, K. B., Schwinger, J., Fassbender, A. J., Landschützer, P., Yamaguchi, R., Frenzel, H., et al. (in review). Seasonal variability of the surface ocean carbon cycle: a synthesis. *Global Biogeochemical Cycles*.

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

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

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

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

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

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

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

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

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

Séférian, R., Nabat, P., Michou, M., Saint-Martin, D., Voldoire, A., Colin, J., et al. (2019). Evaluation of CNRM Earth System Model, CNRM-ESM2-1: Role of Earth System Processes in Present-Day and Future Climate. *Journal of Advances in Modeling Earth Systems*, *11*(12), 4182–4227. https://doi.org/https://doi.org/10.1029/2019MS001791

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

https://doi.org/https://doi.org/10.1029/2008GB003311
Stock, C. A., Dunne, J. P., Fan, S., Ginoux, P., John, J., Krasting, J. P., et al. (2020). Ocean Biogeochemistry in GFDL's Earth System Model 4.1 and Its Response to Increasing Atmospheric CO_2. *Journal of Advances in Modeling Earth Systems*, *12*(10). https://doi.org/https://doi.org/10.1029/2019MS002043
Takahashi, T., Broecker, W. S., & Bainbridge, a. E. (1981). The alkalinity and total carbon dioxide concentration in the world oceans. *Carbon Cycle Modelling, SCOPE*, *16*(3078), 271–286. Retrieved from http://data.ieda.org.cn/upload/fckeditor/SCOPE_16_1.5.07_Takahashi_271-286.pdf
Talley, L. D. (2013). Closure of the Global Overturning Circulation Through the Indian, Pacific, and Southern Oceans: Schematics and Transports. *Oceanography*, *26*, 80–97. https://doi.org/10.5670/oceanog.2013.07
Tanhua, T., Hoppema, M., Jones, E. M., Stöven, T., Hauck, J., Dávila, M. G., et al. (2017). Temporal changes in ventilation and the carbonate system in the Atlantic sector of the Southern Ocean. *Deep Sea Research Part II: Topical Studies in Oceanography*, *138*, 26–38. https://doi.org/https://doi.org/10.1016/j.dsr2.2016.10.004
Terhaar, J., Orr, J. C., Gehlen, M., Ethé, C., & Bopp, L. (2019). Model constraints on the anthropogenic carbon budget of the Arctic Ocean. *Biogeosciences*, *16*(11), 2343–2367. https://doi.org/10.5194/bg-16-2343-2019
Terhaar, J., Orr, J. C., Ethé, C., Regnier, P., & Bopp, L. (2019). Simulated Arctic Ocean Response to Doubling of Riverine Carbon and Nutrient Delivery. *Global Biogeochemical Cycles*, *33*(8), 1048–1070. https://doi.org/https://doi.org/10.1029/2019GB006200
Terhaar, J., Kwiatkowski, L., & Bopp, L. (2020). Emergent constraint on Arctic Ocean acidification in the twenty-first century. *Nature*, *582*(7812), 379–383. https://doi.org/10.1038/s41586-020-2360-3
Terhaar, J., Tanhua, T., Stöven, T., Orr, J. C., & Bopp, L. (2020). Evaluation of Data-Based Estimates of Anthropogenic Carbon in the Arctic Ocean. *Journal of Geophysical Research: Oceans*, *125*(6).
Terhaar, J., Torres, O., Bourgeois, T., & Kwiatkowski, L. (2021). Arctic Ocean acidification over the 21st century co-driven by anthropogenic carbon increases and freshening in the CMIP6 model ensemble. *Biogeosciences*, *18*(6), 2221–2240. https://doi.org/10.5194/bg-18-2221-2021
Terhaar, J., Lauerwald, R., Regnier, P., Gruber, N., & Bopp, L. (2021). Around one third of current Arctic Ocean primary production sustained by rivers and coastal erosion. *Nature Communications*, *12*(1), 169. https://doi.org/10.1038/s41467-020-20470-z
Terhaar, J., Frölicher, T., & Joos, F. (2021). Southern Ocean anthropogenic carbon sink constrained by sea surface salinity. *Science Advances*, *7*, 5964–5992. https://doi.org/10.1126/sciadv.abd5964
Terhaar, J., Frölicher, T. L., & Joos, F. (2022). Observation-constrained estimates of the global ocean carbon sink from Earth System Models. *Biogeosciences*, *19*, 4431–4457. https://doi.org/10.5194/bg-19-4431-2022
Terhaar, J., Frölicher, T. L., & Joos, F. (2023). Ocean acidification in emission-driven temperature stabilization scenarios: the role of TCRE and non-CO_2 greenhouse gases. *Environmental Research Letters*, *18*(2). https://doi.org/10.1088/1748-9326/acaf91
Tsujino, H., Urakawa, S., Nakano, H., Small, R. J., Kim, W. M., Yeager, S. G., et al. (2018). JRA-55 based surface dataset for driving ocean–sea-ice models (JRA55-do). *Ocean*

Modelling, *130*, 79–139. https://doi.org/https://doi.org/10.1016/j.ocemod.2018.07.002

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

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

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

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

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

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

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

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

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

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

Yasunaka, S., Manizza, M., Terhaar, J., Olsen, A., Yamaguchi, R., Landschützer, P., et al. (in review). An assessment of CO_2 uptake in the Arctic Ocean from 1985 to 2018. *Global Biogeochemical Cycles*.

Zeng, J., Nojiri, Y., Landschützer, P., Telszewski, M., & Nakaoka, S. (2014). A Global Surface Ocean fCO2 Climatology Based on a Feed-Forward Neural Network. *Journal of Atmospheric and Oceanic Technology*, *31*(8), 1838–1849. https://doi.org/10.1175/JTECH-D-13-00137.1

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