Research

Author for correspondence:
Judith Hauck
e-mail: judith.hauck@awi.de

THE ROYAL SOCIETY
PUBLISHING

Sparse observations induce large biases in estimates of the global ocean CO2 sink: an ocean model subsampling experiment

Judith Hauck[1], Cara Nissen[1,2], Peter Landschützer[3], Christian Rödenbeck[4], Seth Bushinsky[5] and Are Olsen[6,7]

[1]Alfred-Wegener-Institut Helmholtz-Zentrum für Polar- und Meeresforschung, Bremerhaven, Germany
[2]Department of Atmospheric and Oceanic Sciences and Institute of Arctic and Alpine Research, University of Colorado Boulder, Boulder, CO, USA
[3]Flanders Marine Institute (VLIZ), Ostend, Belgium
[4]Max Planck Institut für Biogeochemie, Jena, Germany
[5]School of Ocean and Earth Science and Technology, University of Hawai'i at Mānoa, Department of Oceanography, Honolulu, HI, USA
[6]Geophysical Institute, University of Bergen, Bergen, Norway
[7]Bjerknes Centre for Climate Research, Bergen, Norway

JH, 0000-0003-4723-9652; CN, 0000-0001-5804-3191;
PL, 0000-0002-7398-3293; CR, 0000-0001-6011-6249;
SB, 0000-0001-5106-4678; AO, 0000-0003-1696-9142

Estimates of ocean CO_2 uptake from global ocean biogeochemistry models and pCO_2-based data products differ substantially, especially in high latitudes and in the trend of the CO_2 uptake since 2000. Here, we assess the effect of data sparsity on two pCO_2-based estimates by subsampling output from a global ocean biogeochemistry model. The estimates of the ocean CO_2 uptake are improved from a sampling scheme that mimics present-day sampling to an ideal sampling scheme with 1000 evenly distributed sites. In particular, insufficient sampling has given rise to strong biases in the trend of the ocean

carbon sink in the pCO_2 products. The overestimation of the CO_2 flux trend by 20–35% globally and 50–130% in the Southern Ocean with the present-day sampling is reduced to less than 15% with the ideal sampling scheme. A substantial overestimation of the decadal variability of the Southern Ocean carbon sink occurs in one product and appears related to a skewed data distribution in pCO_2 space. With the ideal sampling, the bias in the mean CO_2 flux is reduced from 9–12% to 2–9% globally and from 14–26% to 5–17% in the Southern Ocean. On top of that, discrepancies of about $0.4\,\mathrm{PgC\,yr^{-1}}$ (15%) persist due to uncertainties in the gas-exchange calculation.

This article is part of a discussion meeting issue 'Heat and carbon uptake in the Southern Ocean: the state of the art and future priorities'.

1. Introduction

The ocean sequesters a remarkably constant fraction of about 25% of anthropogenic CO_2 emissions each year [1,2]. This is quantified in the annual releases of the Global Carbon Budget (GCB) [2] through simulations with global ocean biogeochemistry models (GOBMs) and data products based on measurements of surface ocean pCO_2 (pCO_2-products). This fraction is consistent with interior ocean biogeochemical observation-based estimates [3], atmospheric inversions [2] and observations of atmospheric O_2 to N_2 ratios [4]. There is agreement between the data classes on a mean ocean carbon uptake of 2–$2.2\,\mathrm{PgC\,yr^{-1}}$ in the 1990s, on a stagnation of the ocean sink in the 1990s, and a reinvigoration since the early 2000s. However, there are discrepancies in the temporal evolution of the ocean sink, in particular between the GOBMs and pCO_2 products [1,2]. The rate of increase 2000–2018 amounts to $0.41 \pm 0.04\,\mathrm{PgC\,yr^{-1}\,decade^{-1}}$ in the GOBMs and to $0.69 \pm 0.14\,\mathrm{PgC\,yr^{-1}\,decade^{-1}}$ in the pCO_2 products (table 1). The discrepancy grows further in time, and is a factor of three for the trend since 2010 [1,2]: The ocean sink increased by $0.9\,\mathrm{PgC\,yr^{-1}\,decade^{-1}}$ since 2010 according to the pCO_2 products, but only by $0.3\,\mathrm{PgC\,yr^{-1}\,decade^{-1}}$ according to the GOBMs [2]. The diverging trends since around 2000 stem from the Southern Ocean and the northern high latitudes [2]. The Southern Ocean also stands out as the region of largest discrepancy in the mean flux [1,2]. This discrepancy, however, lies within the uncertainty of the river flux adjustment [5] and its spatial distribution [6,7] that needs to be accounted for when comparing the pCO_2 products' flux estimate, which includes signals of carbon transport from land to ocean via rivers and associated ocean outgassing, with the GOBMs ocean sink estimate without this signal [1,2].

Global ocean biogeochemistry models are general ocean circulation models with biogeochemical modules. They simulate the CO_2 flux at the air–sea interface and the exchange of carbon between the surface and the deep ocean, with all their seasonal, interannual and longer time-scale variations. They further include a low- to intermediately complex representation of the biological carbon cycle [8]. They are closely tied to recent climate change and variability as they are forced with atmospheric reanalysis fields, such as the Japanese 55-year atmospheric reanalysis (JRA55-do) [9]. They reproduce the spatial and temporal variability of surface ocean pCO_2 observations relatively well on large spatial and annual scales [1]. A direct comparison with ocean interior carbon accumulation 1994–2007 [3], as well as with the best estimate of the mean 1990s ocean carbon uptake as assessed by the Intergovernmental Panel on Climate Change (IPCC) fourth Assessment Report (AR4) [10] indicates that the GOBMs underestimate the mean ocean carbon sink by 0.2–$0.4\,\mathrm{PgC\,yr^{-1}}$ [2]. A similar underestimation of ocean carbon uptake in Earth System Models was related to model biases in surface to deep ocean carbon transport [11–13] and in the surface chemical buffer capacity (Revelle factor) [14,15].

Annually updated estimates of the ocean carbon sink from surface ocean pCO_2 observations have been facilitated by the advent of the quality-controlled Surface Ocean CO_2 Atlas (SOCAT) in 2011 [16,17] and its annual updates since 2015 [18]. The SOCATv2022 release contains 33.7 million pCO_2 observations with an estimated accuracy of better than 5 micro atmospheres. These

Table 1. Decadal trend in the ocean carbon uptake over the period 2000−2018 in the Global Carbon Budget 2022 [2] global ocean biogeochemistry models and surface ocean pCO_2 based data products, and in the FESOM-1.4-REcoM2 version used here. The ensemble mean and standard deviation (s.d.) are also given.

dataset	trend 2000−2018 (PgC yr^{-1} decade^{-1})
global ocean biogeochemistry models used in Global Carbon Budget 2022	
CESM2	0.40
NEMO3.6-PISCESv2-gas (CNRM)	0.44
FESOM2.1-REcoM2	0.45
NEMO-PISCES (IPSL)	0.38
MOM6-COBALT (Princeton)	0.37
MRI-ESM2-1	0.40
MICOM-HAMOCC (NorESM1-OCv1.2)	0.38
NEMO-PlankTOM12	0.50
CESM-ETHZ	0.40
MPIOM-HAMOCC6	0.34
ensemble mean $\pm$ 1 std	0.41 $\pm$ 0.04
[a]FESOM-1.4-REcoM2	0.43
pCO_2 products	
CMEMS-LSCE-FFNN	0.62
JMA-MLR	0.52
LDEO HPD	0.63
[a]MPI-SOM-FFN	0.95
NIES-NN	0.85
OS-ETHZ-GRaCER	0.60
[a]Jena CarboScope	0.67
ensemble mean $\pm$ 1 std	0.69 $\pm$ 0.14

[a]The model and mapping methods used for this subsampling study.

observations are, however, unevenly distributed in space, time and seasons ([18], updated). They cover about 2% of all grid points in an array with monthly 1° longitude × 1° latitude dimensions, 1990 to present. Based on this dataset, mapping methods were developed that can fill the gaps (98%) to obtain full global coverage. The ocean CO_2 uptake can be calculated from these global maps of surface ocean pCO_2 with gas-exchange parametrizations [19,20], observation-derived wind speed data (e.g. from atmospheric reanalysis) and sea surface temperature and salinity. The first methods, based on a data-driven mixed-layer scheme, a neural network, and multi-linear regression were published in 2013/2014 [21–24] and others have followed since then [25–28]. While a first comparison of the different methods identified substantial discrepancies in terms of the amplitude of interannual variations [29], evaluation with independent data indicated that data scarcity presents a larger limitation than potential methodological weaknesses [26]. This is supported by model subsampling experiments that have revealed biases in pCO_2 products due to data scarcity [30–33]. Specifically, insufficient sampling leads to an overestimation of the global and Southern Ocean amplitude of decadal variability by 21% and 31%, respectively [31], and to a 30% overestimation of the mean flux in the Atlantic Ocean [32]. These studies, however, could not assess the confidence in the ocean carbon sink trend estimate since 2000. This is because one study

was based on a large-ensemble Earth System Model testbed of a high emission future scenario [31], which is not comparable to observed atmospheric CO_2 trends in the period 2000–2018, while the other studies were confined to single ocean basins [32,33].

GOBMs were used for many years as the only data class for the ocean carbon sink estimate in the GCB. Ocean carbon sink estimates from pCO_2 products were used as an independent comparison dataset for the GOBMs since 2013 [34]. In the budget 2021, the ocean sink estimate was for the first time obtained as the average of the GOBM ensemble mean (eight GOBMs) and the pCO_2-product ensemble mean (seven pCO_2 products) [2]. Sparsity in surface pCO_2 observations is thus one of the major sources of uncertainties for this combined ocean sink estimate, and particularly for the temporal evolution, including the trend since around 2000. Autonomous observations, e.g. from biogeochemical Argo floats (bgcArgo) [35,36] or uncrewed surface vehicles [37] may potentially close gaps in the observational network [32,33]. This is of particular interest in southern high latitudes that are less regularly accessed by ships, in particular in winter, although the lower accuracy of pCO_2 values derived from float-based pH measurements remains a major challenge [30,35,38].

In this study, we use output from one global ocean biogeochemistry model that contributes to the Global Carbon Budget to assess the effect of data sparsity on the pCO_2-based estimates of the ocean carbon sink. We train two mapping methods with pCO_2 and other environmental data from the GOBM, which was subsampled according to three sampling schemes (SOCAT, SOCAT plus Southern Ocean bgcArgo floats, ideal bgcArgo coverage). Comparing the resulting pCO_2 reconstructions and air–sea CO_2 fluxes to the known model truth, we assess reconstruction biases with respect to the mean flux, its variability and the magnitude of its trend 2000–2018.

2. Methods

(a) Ocean biogeochemistry model simulation

We use the ocean circulation model FESOM1.4 [39] coupled to the ocean biogeochemistry model REcoM2 [1,40,41]. The unstructured mesh has 126 859 surface nodes, roughly equivalent to a $1° \times 1°$ resolution. The surface nodes are unevenly distributed with the lowest resolution in the subtropics, and higher resolution at the equator, the coasts, the southern high and in particular, the northern high latitudes [39]. The model is started from initial conditions (World Ocean Atlas for nutrient fields [42], GLODAPv2 for alkalinity and preindustrial dissolved inorganic carbon [43]). It is spun up from 1850 to 1957 using repeated year atmospheric forcing, taken from the year 1961. The atmospheric forcing fields for the spin-up and for the simulation period 1958–2018 are taken from the JRA55-do Reanalysis Version 1.4.0 [9]. Further, both spin-up and simulation periods are forced with observed atmospheric xCO_2 as provided by the GCB [44], which is the average of atmospheric CO_2 measurements from the Mauna Loa and South Pole stations since 1958 [45,46]. This is converted to pCO_2 using spatiotemporally varying sea-level pressure and the water vapour correction (a function of sea surface temperature and salinity). Carbonate chemistry and air–sea CO_2 exchange are calculated with the mocsy routines [47] that apply a quadratic gas-exchange parameterization [20] (see equations (2.1) and (2.2)). This is the same model version as used in the Global Carbon Budget 2020 [44] and the RECCAP project (https://reccap2-ocean.github.io). The model output was interpolated to a $1° \times 1°$ field using bilinear interpolation.

(b) Sampling masks

We create three sampling masks (figure 1). The first mask is based on SOCATv2019 (gridded version) [18] and thus comparable to the current pCO_2 product submissions to the Global Carbon Budget. A second mask is created from SOCATv2019 and the bgcArgo observations of the Southern Ocean Carbon and Climate Observations and Modeling project (SOCCOM). The SOCCOM floats provide summer and winter data since 2014. Over the entire study period, this mask is dominated by SOCAT, and SOCCOM contributes less than a quarter of Southern Ocean
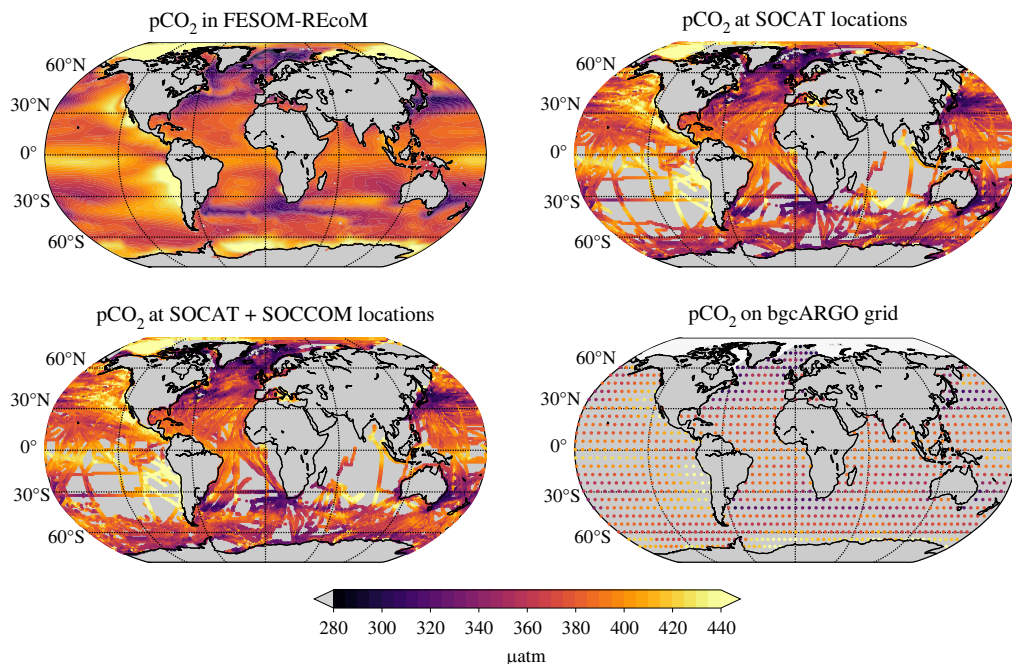
Figure 1. Annual mean partial pressure of CO_2 (pCO_2) averaged over the period 2009–2018 from the output from the global ocean biogeochemistry model FESOM-REcoM. Upper left figure shows the full model output, and the other three panels show the model output after subsampling according to the three masks, as indicated in the title (SOCAT, SOCAT + SOCCOM, bgcArgo). In the bgcArgo panel, the white background shading in the Arctic depicts the Arctic region that is excluded from all analyses. See text for more explanation.

observation since 2014 (figure 2). Finally, the third mask is based on the ideal bgcArgo sampling scheme with 1000 bgcArgo floats which corresponds to a $6° \times 6°$ regular grid [36]. Notably, even though the ideal bgcArgo design aims for full global coverage, the floats do not extend into the Arctic realm following Roemmich *et al.* [36]. We therefore chose to not place hypothetical floats there, using the RECCAP Arctic mask (https://reccap2-ocean.github.io/), which is depicted by lighter background shading in figure 1*d*. In the following, we refer to this ideal sampling grid as bgcArgo. We compare the actual temporally varying data availability from SOCAT and SOCCOM to a hypothetical case with full ideal bgcArgo sampling since 1970. We do not account for movement of floats in the bgcArgo case. We also do not account for higher uncertainty of bgcArgo derived pCO_2 values.

The mean pCO_2 over the period 2009–2018 for the original model output and the three sampling schemes is illustrated in figure 1. The difference in annual mean pCO_2 in the Southern Ocean between full model output and pCO_2 at SOCAT (and SOCAT + SOCCOM) locations can be explained by the summer bias in SOCAT (see electronic supplementary material, figures S1 and S2). The data coverage increases globally from SOCAT to SOCAT + SOCCOM to bgcArgo, except for the Arctic (by design). We therefore exclude the Arctic from our analysis, again using the RECCAP Arctic mask. The number of observations from the three data sources (SOCAT, SOCCOM, bgcArgo) is presented in figure 2. The ideal (hypothetical) bgcArgo grid provides more filled grid cells than SOCAT until 2004 in the global case, until 1994 for the North (north of $30°N$, Arctic excluded), until 2005 in the tropics (and intermittently thereafter, $30°S$ to $30°N$), and always in the Southern Ocean (south of $30°S$). This picture holds throughout seasons, except in the Southern Ocean. Here, SOCAT can offer more summer grid cells covered than bgcArgo since 2006, but only about half the amount of grid cell coverage in winter (electronic supplementary material, Figures S3 and S4). Globally, coverage corresponds to 2.4% of all monthly $1° \times 1°$ grid cells in the
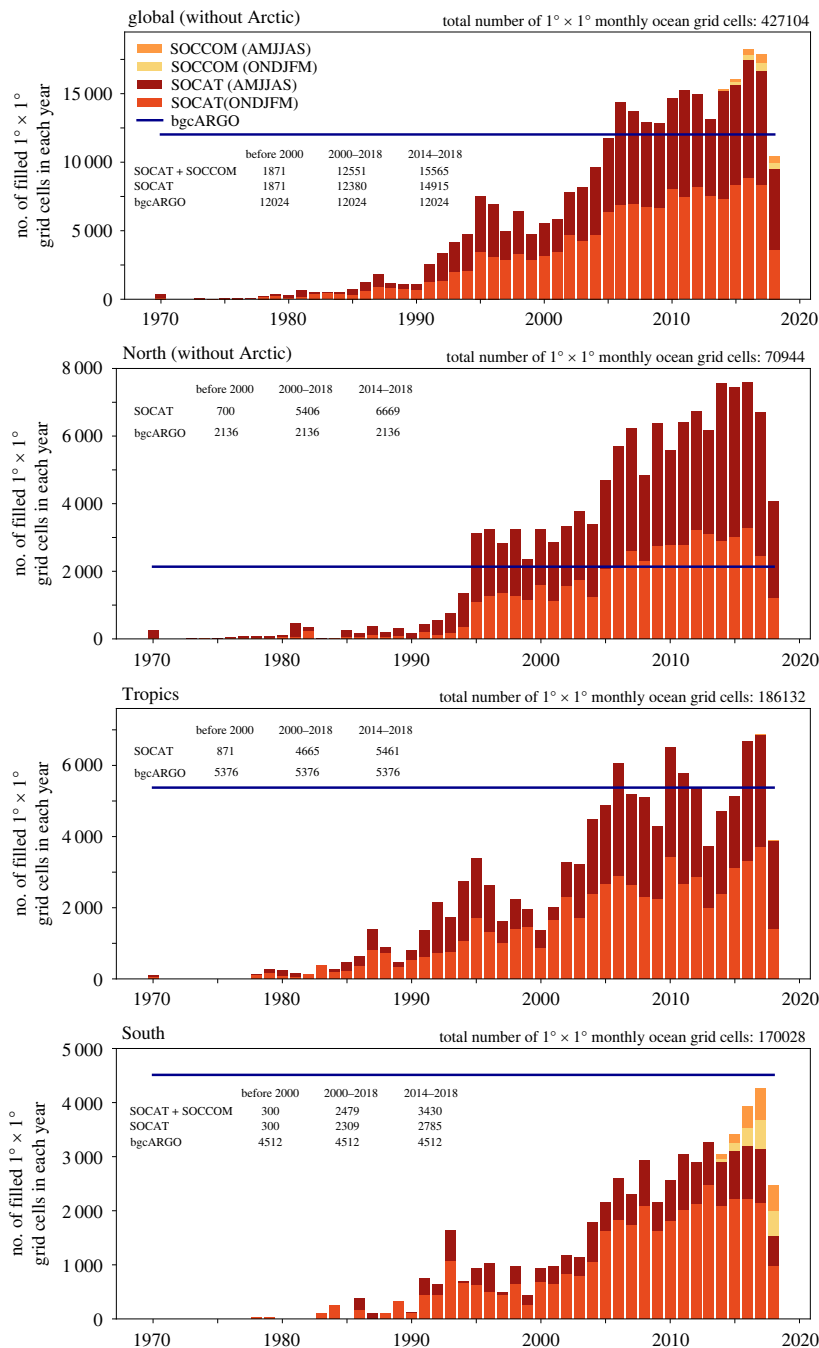
Figure 2. Number of monthly $1° \times 1°$ grid cells covered per year. Shown is the data coverage in the global ocean (top) and three large-scale regions, top to bottom: North (north of 30°N, excluding the Arctic), Tropics (30°S to 30°N), South (south of 30°S). The bars illustrate the data coverage from the SOCAT (dark orange: October to March (Southern Hemisphere summer, Northern Hemisphere winter), dark red: April to September) and SOCCOM (yellow: October to March light orange: April to September) data sets. The hypothetical ideal bgcArgo data coverage is constant in time as indicated by the dark blue horizontal line. Note the different axis scales. Average number of filled grid cells is given within the figures for the periods before 2000, 2000–2018 and 2014–2018. Total number of monthly $1° \times 1°$ ocean grid cells per region is given at the top of the panels.
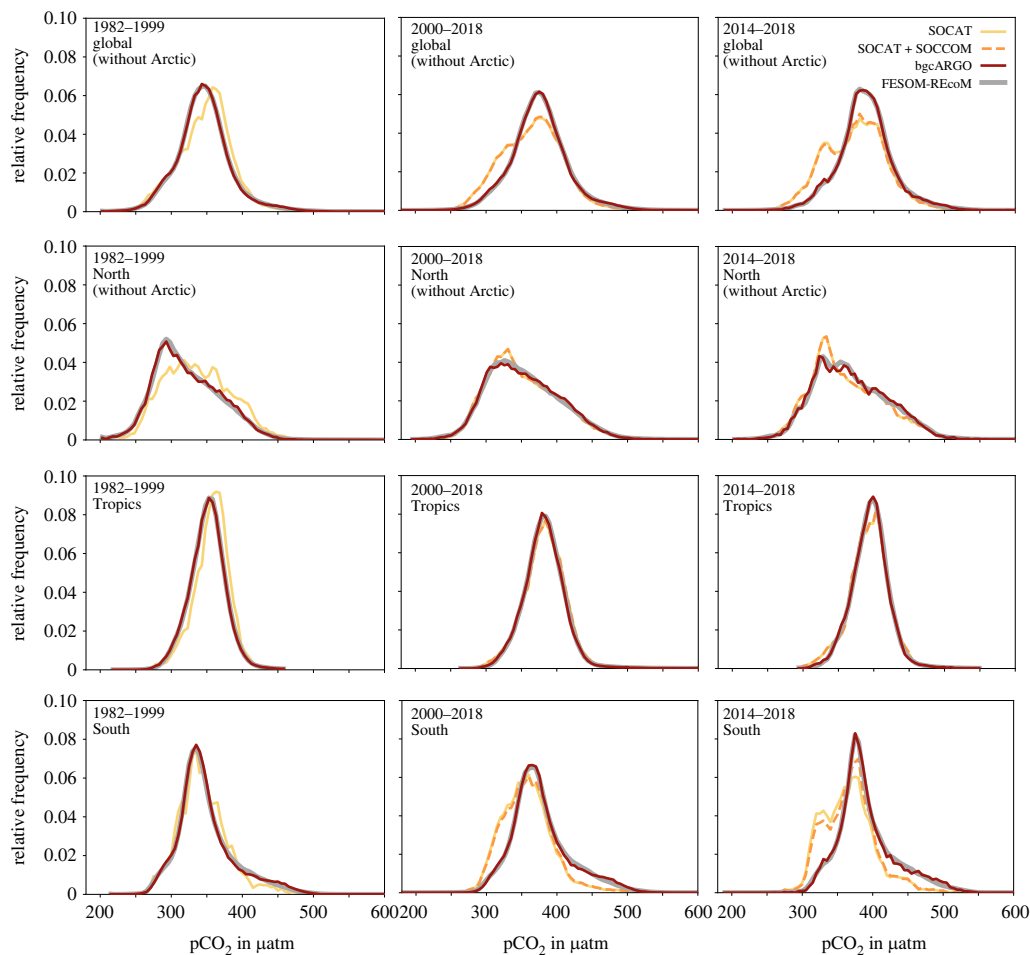
Figure 3. Data distribution in pCO$_2$ space for the regions: (top to bottom) global (no Arctic), North (no Arctic), Tropics, and South, and for the periods (from left to right) 1982–2000, 2000–2018 and the last 5 years 2014–2018. The thick grey line depicts the full model output, and the coloured lines the three sampling masks (SOCAT: yellow; SOCAT + SOCCOM: dashed orange; bgcArgo: red).

bgcArgo scheme, whereas SOCAT would cover 2.6% in the period 2000–2018 and 3.1% in the last five years 2014-2018 (3.3% SOCAT + SOCCOM). The majority of observations come from the North, where 7.6% of the spatio-temporal field 2000–2018 was covered with SOCAT (3.0% in the bgcArgo scheme). In the same period, SOCAT coverage amounts to 2.5% (2.9% bgcArgo) in the tropics, and to 1.4% (2.7% bgcArgo) in the Southern Ocean (figure 2).

The data distribution in pCO$_2$ space also varies between the sampling schemes (figure 3). The SOCAT and SOCAT + SOCCOM sampling masks result in a skewed data distribution relative to the full FESOM-REcoM output. In particular, pCO$_2$ is skewed towards higher pCO$_2$ values before 2000 in all regions, and towards lower pCO$_2$ values after 2000. The latter is most pronounced in the Southern Ocean and hence also in the global mean, but also has a contribution from the north. The bgcArgo scheme captures the data distribution well throughout all time periods and regions.

(c) Mapping methods

We use two mapping methods that were among the first included in the GCB (CarboScope [48], MPI-SOM-FFN [49]). Both methods produce global maps of seawater pCO$_2$ (pCO$_2^{sw}$), and the CO$_2$

flux (FCO_2) is calculated from the gas transfer velocity k_w, the solubility of CO_2 in seawater (K_0) and the difference between pCO_2 in seawater and in the atmosphere (pCO_2^{atm}) [20]:

$$FCO_2 = k_w \cdot K_0 \cdot (pCO_2^{sw} - pCO_2^{atm}). \qquad (2.1)$$

The gas transfer velocity is calculated as a function of wind speed (U), the coefficient of gas transfer (a), the temperature-dependent Schmidt number (Sc) and is scaled with the ice-free area ($1 - F_{ice}$):

$$k_w = a \cdot U^2 \cdot (1 - F_{ice}) \cdot \frac{660^{0.5}}{Sc}. \qquad (2.2)$$

The CarboScope pCO_2 interpolation (also known as Jena mixed layer scheme, Jena-MLS) is a hybrid scheme that uses a multi-linear regression model to link ocean internal carbon sources and sinks to environmental variables, and an explicit interannual correction when pCO_2 observations are available [48]. The following input fields for the regression and for the parametrizations of gas exchange and carbonate chemistry, normally taken from observation-based data sets, are replaced here by FESOM-REcoM output (interpolated from monthly to daily time-steps, and from $1° \times 1°$ to $2.5°$ longitude $\times$ $2°$ latitude): sea surface temperature, sea surface salinity, ice-fraction, alkalinity and mixed layer depth. Wind speed and sea-level pressure are taken from JRA55-do both in FESOM-REcoM and the standard CarboScope pCO_2 interpolation. Usually, CarboScope uses an alkalinity climatology based on sea surface salinity and temperature [50]. In our application, the time-varying FESOM-REcoM alkalinity is used. Likewise, the pCO_2 data are subsampled from the surface pCO_2 field of FESOM-REcoM. In addition, CarboScope uses a prior estimate of the decadal trend of the CO_2 flux taken from an ocean inverse model (OCIM, [51]). In an additional test case, the prior is replaced with the decadal trend of a FESOM-REcoM simulation forced by increasing atmospheric CO_2 and repeated year atmospheric forcing fields (1961, referred to as simulation C in the Global Carbon Budget and RECCAP, no climate change and variability). The scheme thereby attempts to reconstruct the surface pCO_2 fields 1958–2019.

We use two realizations of the CarboScope product. The first realization calculates the air–sea CO_2 flux at the daily time-step used in CarboScope with the gas-transfer velocity from FESOM-REcoM, which was calculated at every 15 min time step and saved as monthly kw660, i.e. normalized to a Schmidt number of 660. The temperature dependence (varying Schmidt number) is taken into account based on the monthly FESOM-REcoM sea surface temperature output. CO_2 solubility is also calculated from FESOM-REcoM sea surface temperature and salinity fields. These daily CO_2 flux fields are then averaged into monthly fields. The gas-transfer velocity in FESOM-REcoM has a mean value of $14.0 \, cm \, h^{-1}$. Atmospheric pCO_2 is also taken from FESOM-REcoM, i.e. from a globally uniform xCO_2 value converted to pCO_2 with spatiotemporally varying sea-level pressure, sea surface temperature and salinity. The second realization uses the native CarboScope gas-exchange formulation based on the quadratic wind speed dependence [52] with the transfer coefficient scaled to match a global mean transfer velocity of $16.5 \, cm \, h^{-1}$ [53] and the spatially resolved atmospheric CO_2 fields from the CarboScope atmospheric inversion [48]. The same wind fields (JRA55-do reanalysis) as in FESOM-REcoM are used. In CarboScope, the surface pCO_2 differs between the two realizations as the ocean internal sources/sinks may respond to the air–sea CO_2 flux.

The MPI-SOM-FFN method uses a self-organizing map (SOM) approach to divide the global ocean into large-scale and highly dynamic biogeochemical provinces, in which pCO_2 is reconstructed with a feed-forward neural network (FFN) model, that has previously been trained with pCO_2 observations and environmental variables [49]. Here, the monthly $1° \times 1°$ MPI-SOM-FFN reconstruction 1982–2018 uses chlorophyll a, sea surface temperature, sea surface salinity, atmospheric CO_2 concentration and mixed layer depth as environmental variables from FESOM-REcoM model output. By training the feed-forward neural network model on pCO_2 data from the FESOM-REcoM model that is subsampled according to the experiment setup of this study, the network reconstructs a nonlinear relationship between environmental drivers and pCO_2 within each biome, that is then used to fill gaps in the sea surface pCO_2.

As for CarboScope, we use two air–sea CO_2 flux datasets for MPI-SOM-FFN that both build on the same monthly $1° \times 1°$ pCO_2 fields, and on FESOM-REcoM output of atmospheric pCO_2. Note that MPI-SOM-FFN has no feedbacks between CO_2 flux and pCO_2, in contrast to CarboScope. Firstly, CO_2 flux is calculated with FESOM-REcoM output (kw660, sea surface temperature, sea surface salinity) according to equations (2.1) and (2.2) and we thereby take into account the temperature dependency of the gas-transfer velocity as in the model equations [20,47]. Secondly, following the native MPI-SOM-FFN methodology, the air–sea CO_2 exchange is calculated with a quadratic gas-exchange velocity parametrization [52], scaled to a global mean transfer velocity $\langle k_w \rangle$ of $16.5\,cm\,h^{-1}$. The solubility of CO_2 in seawater is calculated from sea surface temperature and salinity output from FESOM-REcoM and the formulation of Weiss [54]. The transfer velocity k_w is calculated from ERA5 6-hourly u- and v-wind components from which the square of wind speed (i.e. U^2) is calculated at 6-hourly temporal resolution. The second moment wind speed is then obtained by averaging into monthly means ($\langle U^2 \rangle$) before calculating gas-exchange according to equations (2.1) and (2.2). Additionally, the air–sea CO_2 exchange is limited to ice-free ocean areas, determined by the sea-ice fraction provided by the FESOM-REcoM model output.

(d) Statistics

We primarily compare the mean CO_2 flux estimate over the last decade of the reconstructions (2009–2018, $FCO_2^{2009-2018}$) and the trend 2000–2018 which is calculated as a linear fit over these years. We further calculate the standard deviation and correlation coefficient of the detrended CO_2 flux time-series 1982–2018 as a measure of the amplitude and phasing of interannual variability of the large-scale fluxes [2,29], acknowledging that the so-calculated metrics also contain signals from decadal variability.

3. Results

The results section is structured into three parts. Firstly, we analyse the air–sea CO_2 fluxes and the mismatch between pCO_2 products and known model truth based on the air–sea CO_2 fluxes calculated from mapped pCO_2 and the FESOM-REcoM gas-exchange formulation (§3a). This is to analyse the effect of sampling distribution without interference of gas-exchange calculations. Secondly, we compare differences in reconstructed surface ocean pCO_2 (§3b). Finally, we analyse the difference between using the ocean model's or the pCO_2 products' native gas-exchange formulation for the calculation of air–sea CO_2 flux from mapped pCO_2 (§3c).

(a) Air–sea CO_2 fluxes

The reconstructed global and regional CO_2 flux time-series from the three sampling experiments with the same gas-exchange calculation as in FESOM-REcoM are presented in figure 4, and the key statistics of the global CO_2 flux in figure 5. Both mapping methods overestimate the mean CO_2 uptake 2009–2018 and the trend 2000–2018 in the SOCAT sampling scheme. In the MPI-SOM-FFN method, the 12% overestimation of the mean in the SOCAT scheme is reduced to 9% in bgcArgo. The 9% overestimation in CarboScope (SOCAT) vanishes in the bgcArgo scheme (2% underestimation, figures 4 and 5). For the trend since 2000, the overestimation of 35% in the SOCAT scheme turns into a 10% underestimation with the bgcArgo scheme for MPI-SOM-FFN and the 20% overestimation to a 2% underestimation in CarboScope. Increasing data availability from SOCAT to the bgcArgo scheme hence leads to substantially improved agreement on the trend since 2000 (FESOM-REcoM: $0.43\,PgC\,yr^{-1}\,decade^{-1}$, MPI-SOM-FFN (SOCAT): $0.58\,PgC\,yr^{-1}\,decade^{-1}$, (bgcArgo): $0.39\,PgC\,yr^{-1}\,decade^{-1}$, CarboScope (SOCAT): $0.52\,PgC\,yr^{-1}\,decade^{-1}$, (bgcArgo): $0.43\,PgC\,yr^{-1}\,decade^{-1}$), and on the mean flux 2009–2018 in both products (FESOM-REcoM: $2.44\,PgC\,yr^{-1}$, MPI-SOM-FFN (SOCAT): $2.73\,PgC\,yr^{-1}$, (bgcArgo): $2.67\,PgC\,yr^{-1}$, CarboScope (SOCAT): $2.66\,PgC\,yr^{-1}$, (bgcArgo): $2.39\,PgC\,yr^{-1}$). The difference between the reconstructions based on SOCAT and SOCAT + SOCCOM sampling
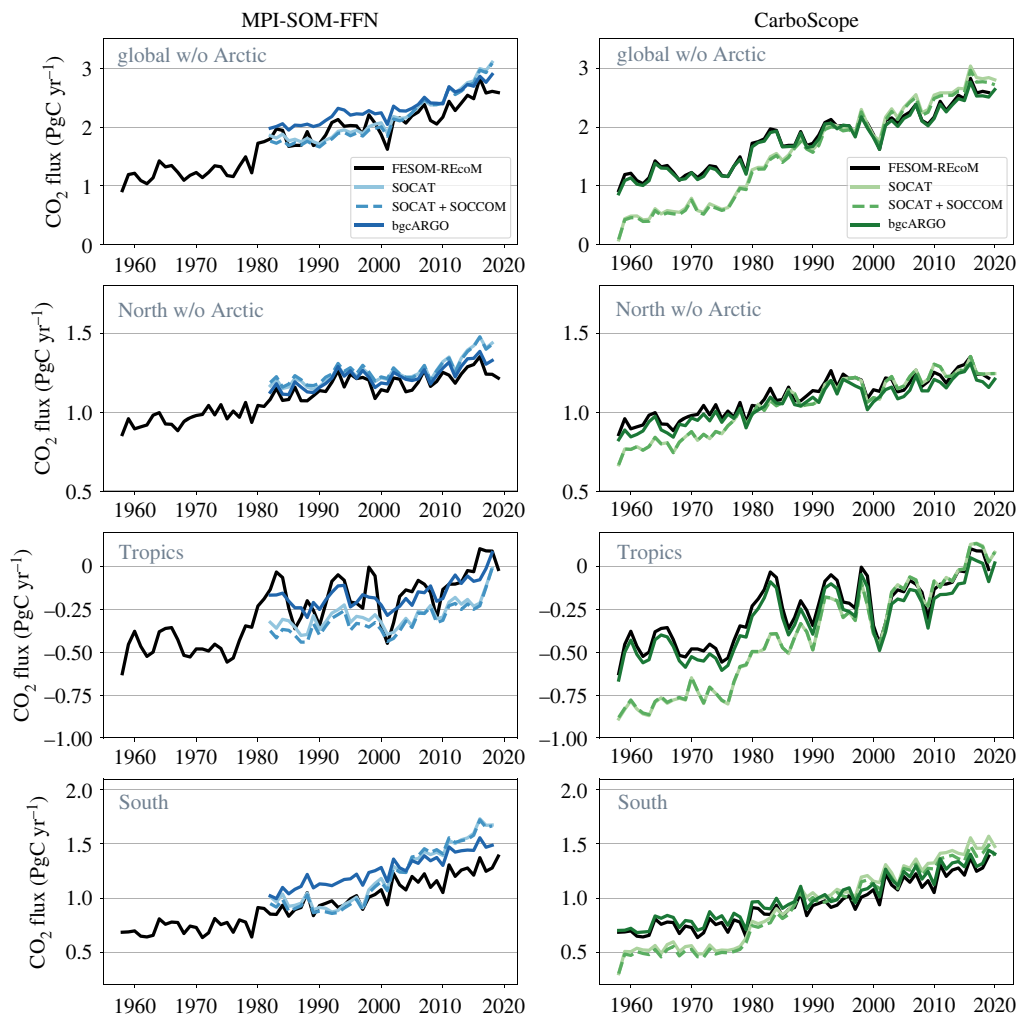
Figure 4. Annual mean air-to-sea CO_2 flux (PgC yr^{-1}) from the pCO_2 reconstructions (coloured lines) compared with the original FESOM-REcoM CO_2 flux (black line). The left column shows the MPI-SOM-FFN and the right column the CarboScope reconstructions for the three sampling schemes as indicated in the figures. MPI-SOM-FFN and CarboScope fluxes are calculated from mapped pCO_2 and FESOM-REcoM gas-transfer velocity. From top to bottom: Global without Arctic, North (north of 30°N) with Arctic excluded, Tropics (30°S to 30°N), South (south of 30°S). Positive fluxes denote a flux into the ocean.

schemes is small (global and Southern Ocean $FCO_2^{2009-2018}$ 0.04–0.05 PgC yr^{-1}). The amplitude of variability is with 0.14 PgC yr^{-1} relatively well captured by MPI-SOM-FFN (SOCAT) in comparison to FESOM-REcoM (0.16 PgC yr^{-1}) and is somewhat underestimated in the bgcArgo sampling scheme (0.10 PgC yr^{-1}). The amplitude of variability of CarboScope turns from a slight underestimation with SOCAT sampling (0.14 PgC yr^{-1}) to a good match with bgcArgo sampling (0.16 PgC yr^{-1}). The CarboScope method reproduces the phasing of temporal variability, as measured by the correlation coefficient with the FESOM-REcoM annual time-series, more skilfully than MPI-SOM-FFN, and responds more strongly to higher and more evenly distributed data availability (MPI-SOM-FFN (SOCAT): 0.69, (bgcArgo): 0.83, CarboScope (SOCAT): 0.81, (bgcArgo): 0.99).

A notable result is found in the Southern Ocean (figure 4). Here, the overestimation of the trend with the SOCAT sampling scheme of 50% in CarboScope and 130% in MPI-SOM-FFN is eliminated with the bgcArgo scheme (trend 2000–2018 FESOM-REcoM: 0.13 PgC yr^{-1} decade^{-1},
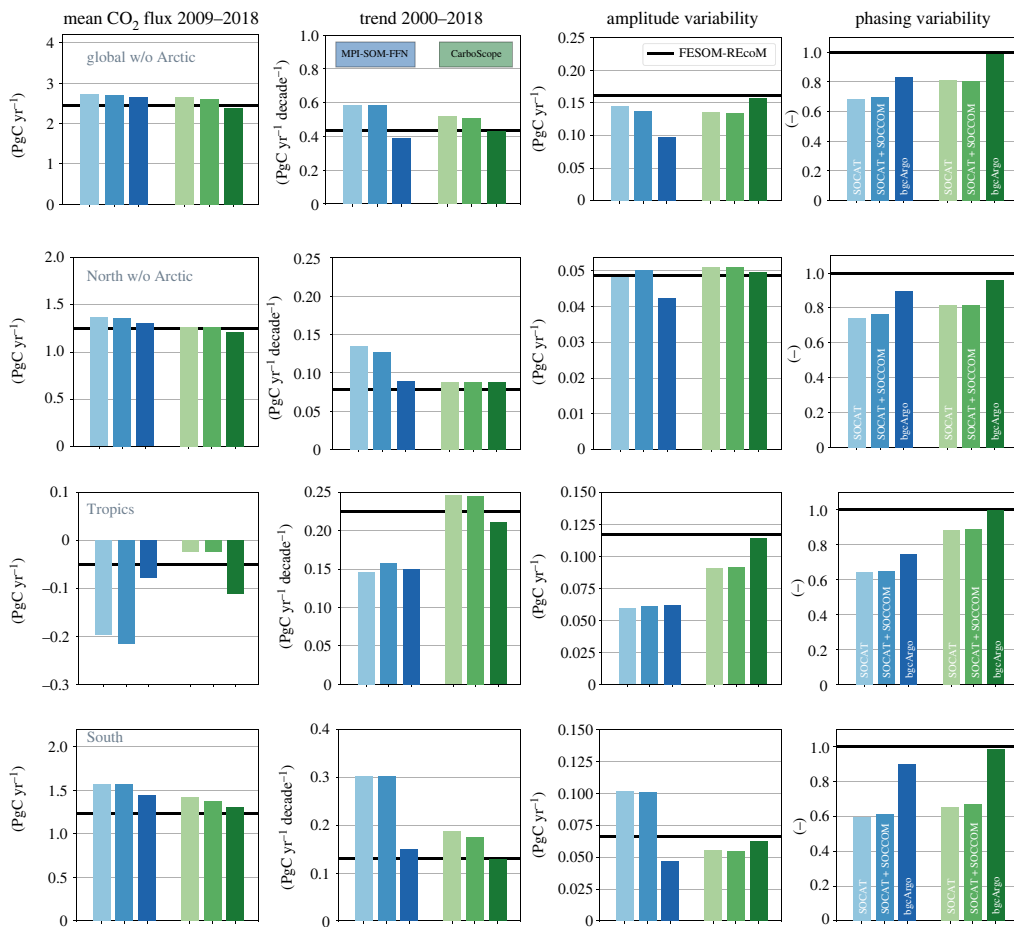
Figure 5. Effect of sampling distribution on statistics of annual mean air-to-sea CO_2 flux (PgC yr^{-1}) calculated from the pCO$_2$ reconstructions with FESOM-REcoM gas-transfer velocity (coloured bars) compared with the original FESOM-REcoM CO_2 flux (black line). From left to right: Mean CO_2 flux 2009–2018 (PgC yr^{-1}), trend 2000–2018 (PgC yr^{-1} decade^{-1}), amplitude of variability (standard deviation of detrended time-series) 1982–2018, phasing of variability (correlation coefficient of detrended time-series) 1982–2018. The blue bars show the MPI-SOM-FFN and the green bars the CarboScope reconstructions for the three sampling schemes as indicated in the figures (SOCAT: light colour, SOCAT + SOCCOM: medium colour, bgcArgo: dark colour). These statistics are calculated from the area-integrated time-series for (top to bottom): global excluding the Arctic, North excluding the Arctic, Tropics, South.

MPI-SOM-FFN SOCAT: 0.30 PgC yr^{-1} decade^{-1}, bgcArgo: 0.15 PgC yr^{-1} decade^{-1}; CarboScope (SOCAT): 0.19 PgC yr^{-1} decade^{-1}, bgcArgo: 0.13 PgC yr^{-1} decade^{-1}). Similarly, the amplitude of variability is reduced by a factor of two in MPI-SOM-FFN in the bgcArgo relative to SOCAT sampling scheme. The correlation with the FESOM-REcoM time-series increases in both products with the bgcArgo scheme (MPI-SOM-FFN: from 0.59 to 0.90, CarboScope: from 0.65 to 0.98). Interestingly, an erroneous acceleration of the Southern Ocean carbon sink since the late 1990s in MPI-SOM-FFN is corrected by increased data availability. The mean flux in the bgcArgo scheme is then, however, consistently overestimated by about 0.2 PgC yr^{-1} (16%, 2009–2018) in MPI-SOM-FFN. The temporal evolution of CarboScope after 1980 is not as strongly affected as in MPI-SOM-FFN, although the mean FCO$_2^{2009-2018}$ and particularly the trend are also reduced in CarboScope. The bgcArgo sampling scheme corrects a low bias in the CarboScope CO_2 flux prior to 1985.

An improvement of the reconstructions with increased and more evenly distributed data availability also occurs in the tropics (figure 4). In the bgcArgo scheme, the mean $FCO_2^{2009-2018}$ is well captured with a slight overestimation of the ocean carbon outgassing by both products (FESOM-REcoM: $-0.05\,\mathrm{PgC\,yr^{-1}}$, MPI-SOM-FFN: $-0.08\,\mathrm{PgC\,yr^{-1}}$, CarboScope: $-0.11\,\mathrm{PgC\,yr^{-1}}$). This is a substantial improvement compared with the SOCAT scheme in MPI-SOM-FFN ($-0.20\,\mathrm{PgC\,yr^{-1}}$), whereas the offset is somewhat larger and of opposite sign for CarboScope in the bgcArgo sampling scheme (SOCAT, $-0.02\,\mathrm{PgC\,yr^{-1}}$). The phasing of temporal variability improves in the bgcArgo scheme (correlation coefficients MPI-SOM-FFN: from 0.64 to 0.75, CarboScope: from 0.89 to 1.00). The amplitude of variability is too low in both products, but the discrepancy is eliminated in CarboScope with the bgcArgo scheme. The availability of data prior to 1990 in the bgcArgo scheme corrects a large tropical outgassing signal in CarboScope based on SOCAT (and SOCCOM, figure 4), which we relate to the very low data availability before 1990 and to the skewed SOCAT pCO_2 distribution towards higher pCO_2 (figure 3, tropics, 1982–1999).

In the North (north of $30°$N, Arctic excluded), there is little improvement from SOCAT to ideal bgcArgo sampling for the mean CO_2 uptake and amplitude of variability (figures 4 and 5). This is because more observed grid cells have been available in SOCAT since 1995 than in bgcArgo (figure 2). The reconstructed mean flux in MPI-SOM-FFN bgcArgo scheme is 4% higher than in the FESOM-REcoM data. Interestingly, the effect of the skewed SOCAT distribution (figure 3) on the CO_2 flux in the North is small, but notable for the last 5 years and the trend in MPI-SOM-FFN. The overestimation of the trend (2000–2018) is reduced in MPI-SOM-FFN with the ideal sampling (FESOM-REcoM: $0.08\,\mathrm{PgC\,yr^{-1}\,decade^{-1}}$, MPI-SOM-FFN (SOCAT): $0.14\,\mathrm{PgC\,yr^{-1}\,decade^{-1}}$, MPI-SOM-FFN (bgcArgo): $0.09\,\mathrm{PgC\,yr^{-1}\,decade^{-1}}$; no change in CarboScope: $0.09\,\mathrm{PgC\,yr^{-1}\,decade^{-1}}$) and the correlation is increased for MPI-SOM-FFN and CarboScope (MPI-SOM-FFN: from 0.74 to 0.90, CarboScope: from 0.82 to 0.96). As in the other regions, an underestimation of the CarboScope CO_2 flux before 1980 in the SOCAT scheme is corrected with the bgcArgo scheme.

The spatial patterns of the air–sea CO_2 fluxes averaged over the period 2009–2018 are very well reproduced by the MPI-SOM-FFN and CarboScope methods in all sampling schemes (figure 6). The only exception is anomalous outgassing in the Arctic in the bgcArgo scheme, which does not include any input data from this region. In MPI-SOM-FFN, the biases go generally into the same direction as the fluxes, i.e. fluxes into the ocean are overestimated in magnitude and fluxes out of the ocean are equally overestimated. Exceptions are a smaller outgassing signal along the west coast of South America and smaller uptake in the south Atlantic high uptake regions. A higher sampling density reduces the overestimation throughout the ocean, but reduces the overestimated outgassing more than the overestimated uptake. In CarboScope, biases are smaller but with the same patterns. However, the bias in the tropical outgassing regions is negligible, leaving the biases mostly located in high-wind CO_2 uptake regions. In the bgcArgo scheme, patches of over- and underestimation alternate. We next inspect how well surface pCO_2 is reconstructed, before investigating the role of transfer velocity in §3c.

(b) Surface pCO$_2$

We inspect area-weighted pCO_2 biases as the differences between regionally averaged (area-weighted) annual time series of the pCO_2 reconstructions and the original FESOM-REcoM model output (figure 7), as well as spatial patterns of the 2009–2018 mean pCO_2 bias (figure 8). We note that regionally averaged pCO_2 biases are within -10 to $+10\,\mu$atm outside the Arctic (figure 7). Locally, the biases can be larger (figure 8).

In all regions, the interannual variability of the bias remains unimproved from SOCAT to bgcArgo sampling distribution for MPI-SOM-FFN (figure 7). In CarboScope, the interannual varying component of the bias is reduced with improved data coverage. In the tropics, the pCO_2 reconstruction is improved in the bgcArgo compared with the SOCAT sampling scheme in terms of the mean pCO_2 (both products), and variability (CarboScope; figure 7). A larger positive pCO_2 bias before 1985 and a small negative bias after 2000 in CarboScope (SOCAT) vanishes in the
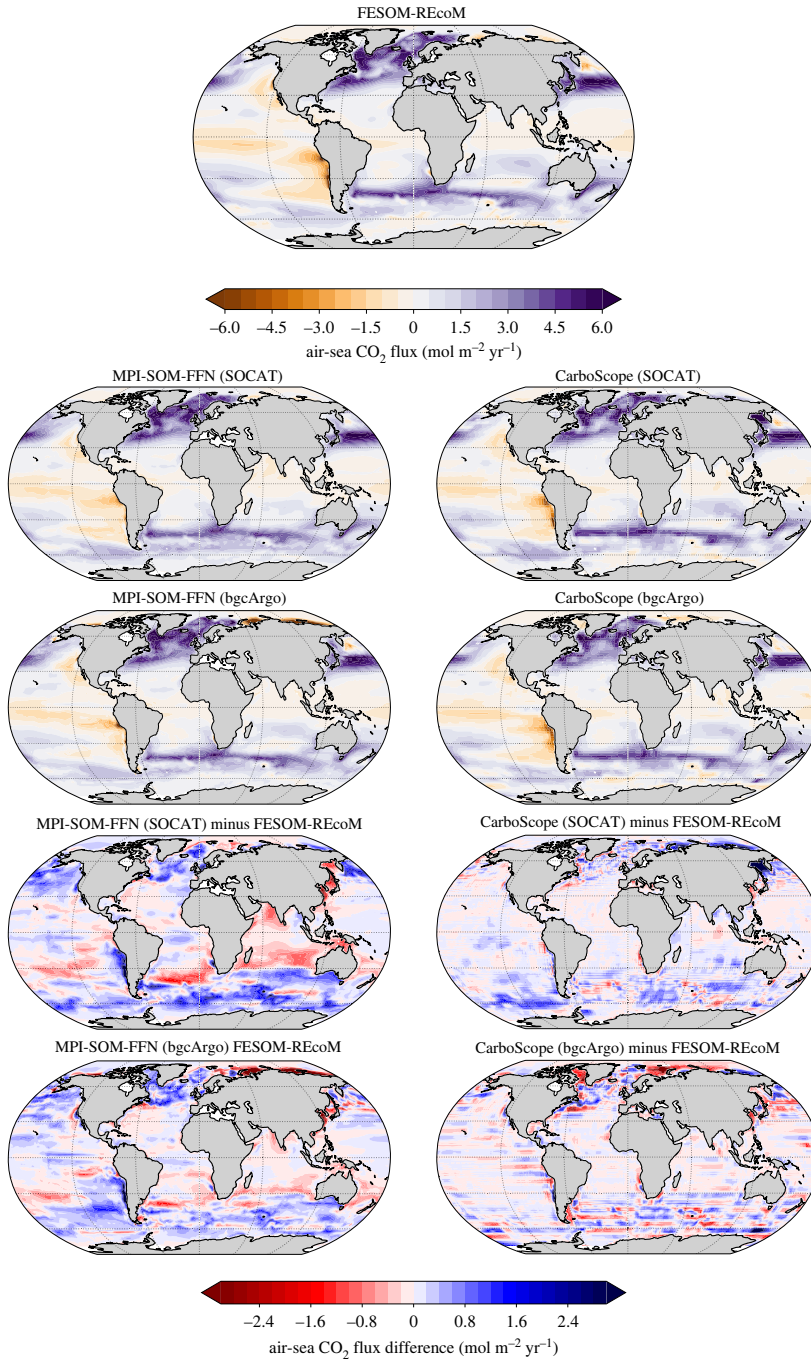
Figure 6. Air−sea CO$_2$ flux in FESOM-REcoM and in the reconstructions by MPI-SOM-FFN (left) and CarboScope (right) products. Positive fluxes (purple) denote a flux into the ocean. In the difference maps, positive numbers (blue) denote a larger flux into the ocean (or a smaller flux out of the ocean) in the reconstruction than in FESOM-REcoM. Top: FESOM-REcoM, then from top to bottom: reconstructed air−sea CO$_2$ flux in SOCAT sampling scheme, in bgcArgo scheme, and difference in air−sea CO$_2$ flux between reconstruction and FESOM-REcoM in SOCAT and bgcArgo schemes, as indicated in the titles. Only the SOCAT and bgcArgo sampling schemes are shown. The differences between SOCAT and SOCAT + SOCCOM sampling schemes are small.
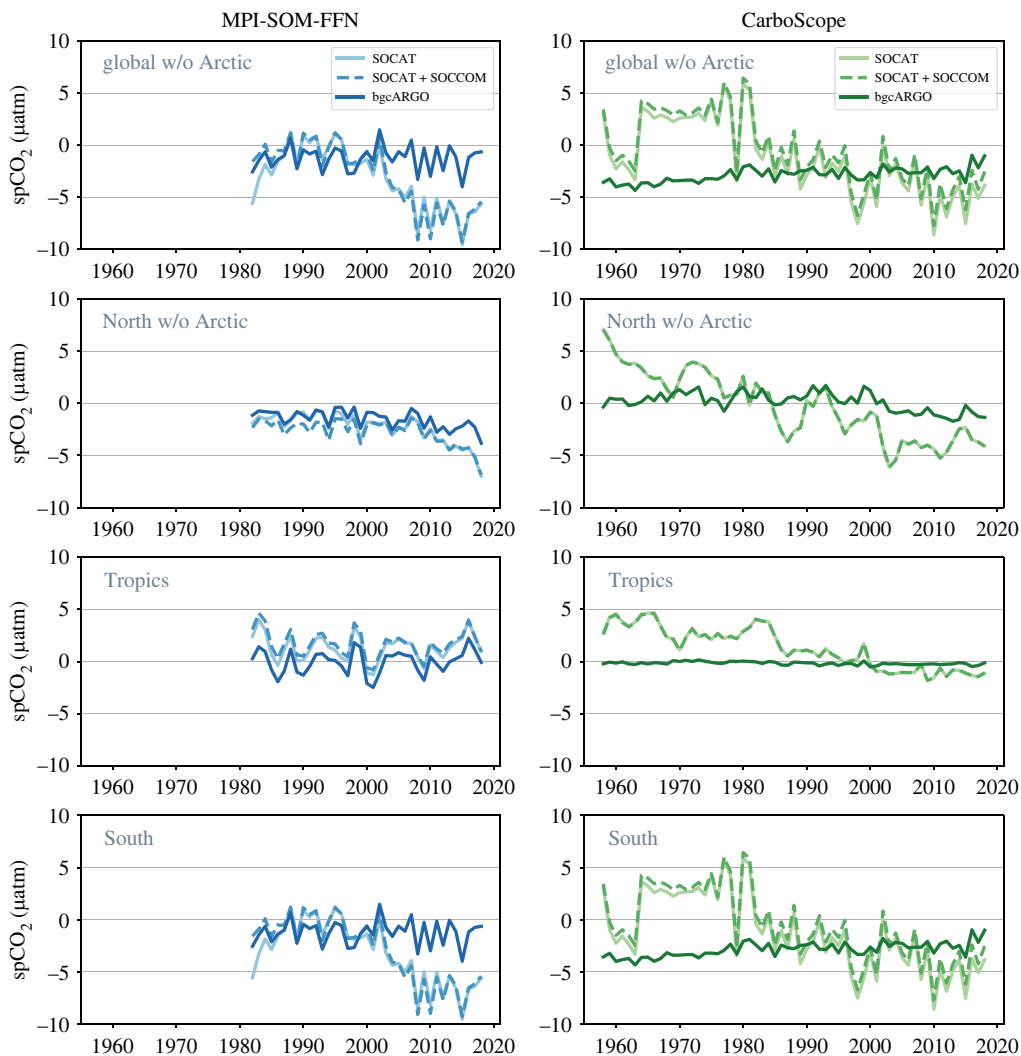
Figure 7. Biases in reconstructed annual mean surface pCO_2 (μatm) (coloured lines) as calculated from the two pCO_2 products minus the original FESOM-REcoM pCO_2. The left column shows the MPI-SOM-FFN and the right column the CarboScope reconstructions for the three sampling schemes as indicated in the figures. From top to bottom: Global (excluding Arctic), North (excluding Arctic), Tropics, South.

bgcArgo case. Thus, pCO_2 is skilfully reproduced in the tropics with data availability as in the hypothetical bgcArgo sampling scheme.

In the North and the South, pCO_2 is underestimated by both products in the regionally averaged time-series. In the North, the mean bias as well as a trend towards increasingly negative bias since 2005 is reduced in the bgcArgo relative to the SOCAT sampling scheme (both products). However, a bias in the mean of about -2 μatm persists in MPI-SOM-FFN and the bias varies between about -2 and $+2$ μatm in CarboScope. A similar result is found in the South, where the SOCAT (and SOCAT + SOCCOM) sampling scheme leads to an increasingly negative pCO_2 bias between 2002 and 2008 in MPI-SOM-FFN (from near 0 to -8 μatm), which persists thereafter. This bias in the temporal evolution vanishes with more data availability in the bgcArgo case. The SOCCOM floats do not reduce this bias, which is located mostly in the Weddell Sea and coastal regions (figure 8). The Southern Ocean mean pCO_2 bias is negative in the bgcArgo scheme,
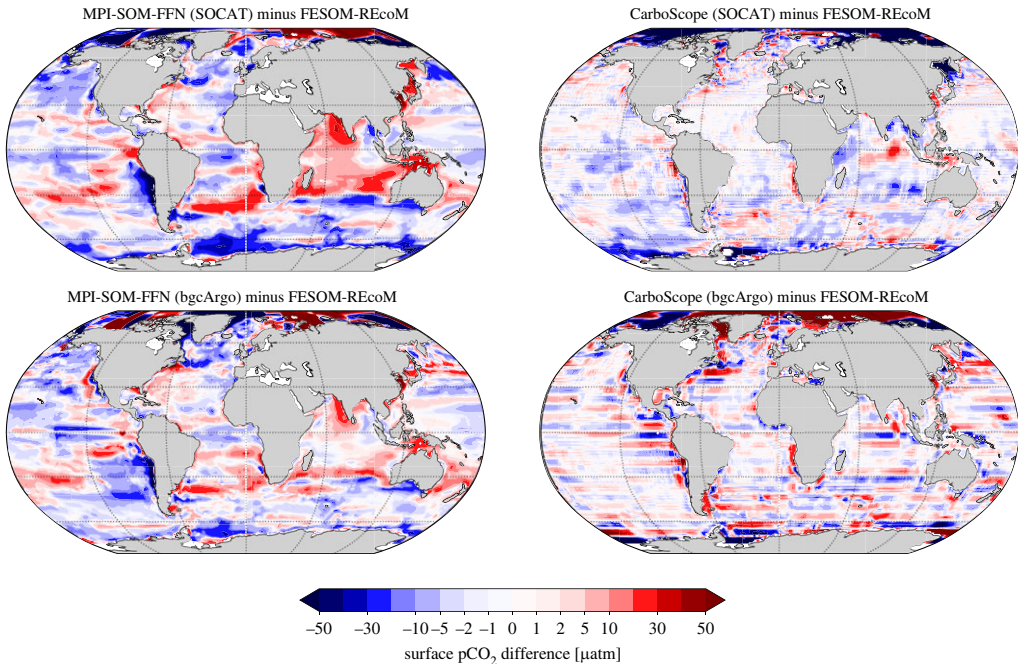
Figure 8. Spatial patterns of biases in reconstructed surface pCO_2 (average 2009–2018) as calculated from the two pCO_2 products minus the original FESOM-REcoM pCO_2, for MPI-SOM-FFN (left) and CarboScope (right) in the sampling schemes SOCAT (top) and bgcArgo (bottom).

varying between −4 and +1 µatm in MPI-SOM-FFN and between −4 and −1 µatm in CarboScope (figure 7).

The maps of the mean pCO_2 bias (figure 8) illustrate that the pCO_2 biases are only partly co-located with the CO_2 flux biases (figure 6). In fact, CarboScope pCO_2 biases (outside of the Arctic) stem mostly from the Weddell and Ross Seas and a small region in the northwest Pacific. These Southern Ocean regions are usually ice-covered and it is unlikely that pCO_2 biases in these regions lead to substantial biases in CO_2 flux. In MPI-SOM-FFN, the negative pCO_2 biases are more equally distributed in the Southern Ocean, but are also not co-located with the CO_2 flux biases (figures 6 and 8).

In summary, we conclude from the analysis of CO_2 flux and pCO_2 biases that more and regularly spaced observations can cure or substantially reduce biases in the temporal evolution of the observation-derived CO_2 flux estimates. However, even with regular $6° \times 6°$ sampling schemes, a discrepancy in the mean CO_2 flux between FESOM-REcoM and MPI-SOM-FFN persists in the Southern Ocean, which suggests a methodological origin (figure 4). The pCO_2 comparison (figures 7 and 8) suggests that the generally larger CO_2 flux in the pCO_2 products with near perfect sampling (bgcArgo) may be related to a persistent underestimation of pCO_2, but with an additional imprint of high wind speed regions. The comparisons so far were conducted with the reconstructed pCO_2 fields and a consistent use of the gas-exchange calculation as in FESOM-REcoM. Next, we investigate additional discrepancies in the mean CO_2 flux, when using the method's native gas-exchange calculations.

(c) Gas-exchange calculation

In addition to artefacts due to skewed data distribution, choices in gas-exchange calculations contribute to discrepancies between air–sea CO_2 fluxes from mapping methods and from the

ocean biogeochemistry model. The difference between air-sea CO_2 flux calculated with FESOM-REcoM gas transfer velocity (as used for figures 4–6) and the MPI-SOM-FFN and CarboScope native gas-exchange calculation is not only substantial in magnitude, but also increases over time and affects the amplitude of interannual variations (figure 9). For example, the difference between the two gas-exchange calculations nearly doubles between 1982 and 2018 from $0.18\,\mathrm{PgC\,yr^{-1}}$ to $0.34\,\mathrm{PgC\,yr^{-1}}$ for the global flux in MPI-SOM-FFN, with roughly equal contributions from the North and South. The air–sea CO_2 flux in the tropics is hardly affected. Similarly, the discrepancy between the two ways to calculate gas-exchange increases by more than a factor of five from $0.07\,\mathrm{PgC\,yr^{-1}}$ to $0.42\,\mathrm{PgC\,yr^{-1}}$ for the global flux in CarboScope. Here, the offset is larger in the North ($0.13\,\mathrm{PgC\,yr^{-1}}$) than in the South ($0.06\,\mathrm{PgC\,yr^{-1}}$) at the beginning of the time-series in 1958, but reaches similar levels towards the end of the time-series (around $0.22\,\mathrm{PgC\,yr^{-1}}$). Note that the native gas-exchange experiment with CarboScope was conducted with a climatological FESOM prior rather than the OCIM prior, but the difference related to the choice of prior is small in the ideal sampling case (compare green and grey dotted lines in figure 9). Choice of prior affects mostly the period before 1990 in the SOCAT and SOCAT + SOCCOM sampling cases (see electronic supplementary material, figure S6). As a result, the amplitude of interannual variations is enlarged with the product's native gas-exchange formulations, as is the trend since 2000 (figure 9, see also electronic supplementary material, figure S7 for the barplot equivalent to figure 5 but with MPI-SOM-FFN native gas-exchange calculation).

We further use the gridded FESOM-REcoM output to test the effects of offline gas-exchange calculation with monthly time stepping as is done in the MPI-SOM-FFN and most other products (CarboScope uses a daily time-step), and choice of the gas transfer coefficient (figure 9 third column). Differences between the direct model output (FESOM-REcoM, online calculation at every 15-min time-step) and the CO_2 flux recalculated from monthly averaged pCO_2 and gas-transfer velocity fields are on the order of 0.07 to $0.09\,\mathrm{PgC\,yr^{-1}}$ in the Southern Ocean and even smaller elsewhere (compare green and black line in figure 9 third column). We interpret this difference as an averaging bias resulting from the high winds in the Southern Ocean whose effect is overestimated when the gas-exchange calculation is done on a monthly time-step. An interpolation bias in pCO_2 and auxiliary fields may contribute as well (interpolation bias in CO_2 flux is small, $0.03\,\mathrm{PgC\,yr^{-1}}$). Tests with 6-hourly model output result in an estimate of the error associated with using monthly time-steps for gas-exchange calculation of 0.05–$0.06\,\mathrm{PgC\,yr^{-1}}$ in the Southern Ocean and globally (as biases of $-0.03\,\mathrm{PgC\,yr^{-1}}$ in the tropics and $+0.03\,\mathrm{PgC\,yr^{-1}}$ in the North compensate). We note that this error is small, but it has the same spatial distribution as the CO_2 flux bias (compare electronic supplementary material, figure S5 and figure 6) leading to CO_2 uptake that is too strong in the temperate and high latitudes. We conclude that the procedure of using monthly fields to calculate gas-exchange works within reasonably small uncertainty.

Another source of uncertainty is the recommended and commonly applied scaling of the coefficient of gas-transfer a to reach a global mean value of the gas-transfer velocity of $16.5\,\mathrm{cm\,h^{-1}}$ [53]. This scaling is applied in the native gas-exchange calculations in MPI-SOM-FFN and CarboScope. The coefficient of gas-transfer in FESOM-REcoM is not scaled and the gas-transfer velocity yields a long-term mean of $14.0\,\mathrm{cm\,h^{-1}}$. We test the effect of the scaling of the gas-exchange, which we approximate by scaling the gas-transfer velocity k_w from FESOM-REcoM to $16.5\,\mathrm{cm\,h^{-1}}$ [53], rather than adjusting the coefficient of gas-transfer as commonly done [53,55]. This approach leads to a global flux increase by 0.17–$0.48\,\mathrm{PgC\,yr^{-1}}$, and this difference also increases over time and stems in roughly equal amounts from the North and South (compare green and orange lines in figure 9 third column). A previous study [55] reported a difference in CO_2 flux of $0.07\,\mathrm{PgC\,yr^{-1}}$ between scaled and unscaled JRA forcing (note that this is a different version than the JRA55-do data set, which adjusts biases in JRA55 [9], used in FESOM-REcoM).

Further, we test the sensitivity of FESOM-REcoM to a higher value of the coefficient of gas-transfer $a=0.31$, which is taken as a value at the upper end of the previously reported range [52]. We perform a short test simulation 1981–2019, i.e. without a corresponding spin-up, and thus the results should be interpreted with some caution. Nevertheless, this test illustrates the low
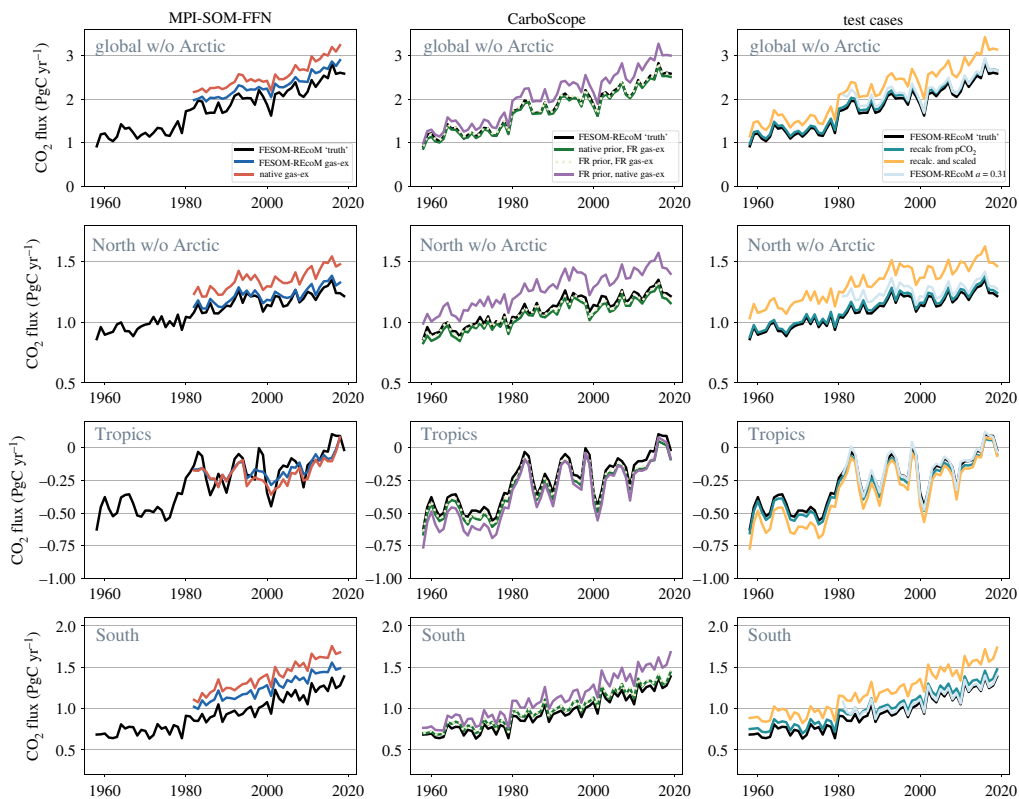
Figure 9. Effect of choices in gas-exchange calculations based on the ideal sampling case for (top to bottom:) global without Arctic, North without Arctic, Tropics, South. All panels show the 'known truth' of the FESOM-REcoM air–sea CO_2 flux (flux integrated on native model mesh, black). Left column: MPI-SOM-FFN CO_2 flux calculated from mapped pCO_2 based on ideal sampling with either FESOM-REcoM piston velocity (blue) or with MPI-SOM-FFN native gas-exchange formulation (red). Middle column: CarboScope CO_2 flux calculated from mapped pCO_2 based on ideal sampling with its native prior from OCIM and FESOM-REcoM piston velocity (green, same as in figure 4), with FESOM-REcoM (FR) prior and FESOM-REcoM piston velocity (light grey, dashed), and with FESOM-REcoM prior and CarboScope native gas-exchange (purple). Right column: Test cases to quantify the error introduced by recalculating the air–sea CO_2 flux from FESOM-REcoM pCO_2 fields (full global coverage) and FESOM-REcoM piston velocity (green), from FESOM-REcoM pCO_2 fields (full global coverage) and FESOM-REcoM gas-transfer velocity scaled to a global mean of 16.5 cm h^{-1} (orange). Also shown is a model experiment with higher gas-transfer coefficient a (0.31 instead of 0.251, light blue).

sensitivity of the ocean biogeochemistry model to choices in the coefficient of gas-transfer. The difference between the two simulations quickly decreases from about 0.25 PgC yr^{-1} globally in the first years to reach 0.07 PgC yr^{-1} at the end of the simulated period (compare black and light blue lines in figure 9 third column). The change over time is likely attributable to the lack of spin-up. This is a notably small effect in CO_2 flux given that the global mean gas-transfer velocity is substantially higher with 17.2 cm h^{-1} relative to the 14.0 cm h^{-1} in the standard simulation. The reason for the low sensitivity of the ocean biogeochemistry model to the gas-transfer coefficient is that the pCO_2 difference between ocean and atmosphere that determines the CO_2 flux between these two compartments quickly adjusts to the initially higher flux [56]. The transfer of carbon between the surface mixed layer and the ocean interior is then the rate-determining step for air–sea CO_2 flux in the model as in the real world. However, the regional distribution of the flux seems to be sensitive to the choice in gas-transfer coefficient, with the difference originating mainly from the North (0.06 PgC yr^{-1}, figure 9).

All in all, the scaling of the gas-transfer coefficient to reach the best estimate of the global gas-transfer velocity explains the majority of the difference between the two products' CO_2 flux reconstructions with their native and FESOM-REcoM gas-transfer velocity. Additional uncertainty comes from the use of a different wind product in the native gas-exchange calculation in MPI-SOM-FFN (ERA5) than in FESOM-REcoM, and a different atmospheric pCO_2 field in the native gas-exchange calculation of CarboScope. Using ERA5 may amount to a difference in the global mean flux of $0.13\,PgC\,yr^{-1}$ based on the assessment of Fay *et al.* [55].

4. Discussion

We investigated the uncertainty in pCO_2 observation based estimates of the ocean carbon sink due to data sparsity. We show that two pCO_2 mapping products using input data according to the SOCAT data distribution overestimate the global mean CO_2 flux, and the trend since 2000 relative to the known model truth (figure 4). Our experiments suggest that the strong decadal variability, and particularly the reinvigoration of the Southern Ocean carbon sink [57] is to a large degree an artefact generated by sparse and decadally skewed data distributions, in line with the finding that the neural network based MPI-SOM-FFN method overestimates decadal variability in the Southern Ocean by 30% [31]. We relate the overestimation of the sink since the late 1990s to the skewed SOCAT data distribution towards low pCO_2 (figure 3) that leads the neural network to also underestimate pCO_2 by 5–10 µatm averaged over the Southern Ocean (figure 7). Interestingly, the dynamics of the reconstructed ocean carbon sink in the 'stagnation' period of the 1990s is less affected by data sparsity and appears to be real. The more robust reconstruction for the 1990s may be related to the distribution of observations in pCO_2 space, which is not skewed towards low pCO_2 values in the period 1982–1999 in our analysis (figure 3).

Even with the ideal sampling scheme, the mean Southern Ocean flux is consistently overestimated by about $0.2\,PgC\,yr^{-1}$ in MPI-SOM-FFN. This seems to be a methodological issue that may be related to the fact that MPI-SOM-FFN aims to minimize the global mismatch with SOCAT observations. While regional pCO_2 biases average out in the first 20 years of the time-series (figure 7), the local pCO_2 biases in high wind speed regions do not compensate in terms of CO_2 flux (figure 6). Further, the biogeochemical provinces that merge regions from different parts of the globe with similar environmental conditions may contribute as well [23]. Southern Ocean mean flux and its trend are also reduced with the ideal sampling in CarboScope, but the effect is smaller than in MPI-SOM-FFN. This is because the CarboScope mixed layer scheme reproduces Southern Ocean pCO_2 well even with the skewed data distribution. A mean pCO_2 offset of -2 to -3 µatm in CarboScope (figure 7) stems from local biases in ice-covered regions which are less important for gas-exchange (figure 8). Similarly, in the North, which is the best observed area, a regularly spaced observation network and equal distribution in pCO_2 space (figure 3) leads to reduction in pCO_2 bias in both products (figure 7) and to a reduced CO_2 flux trend since 2000 in MPI-SOM-FFN. The results illustrate that a process-based mapping method as CarboScope is less prone to biases due to data distribution. This is because a neural network can only reproduce patterns that are included in the training and target data set. With sparse and skewed observations, not all real world pCO_2 values and dependencies on environmental variables are included in the training dataset. CarboScope normally uses a climatological alkalinity field, which is derived from sea surface salinity and temperature [50]. In our study, the model's alkalinity is used as input to CarboScope, and hence additional uncertainty due to the relationship between alkalinity, salinity and temperature and due to missing interannual alkalinity variations in CarboScope are not included in our assessment. The use of data from SOCCOM floats (2014–2018) in addition to SOCAT has a negligible effect on the reconstructed pCO_2 and CO_2 flux (figures 4 and 5). However, an idealized sampling scheme with 1000 regularly ($6° \times 6°$) spaced sampling sites [36] with high-accuracy pCO_2 measurements would be sufficient to reduce the sampling and methodological bias of the mean CO_2 flux to 2–9% globally, and to 5–17% in the Southern Ocean. This would also lead to a reasonably good estimate of the amplitude and particularly the phasing of CO_2 flux variability (figure 5).

Despite the good reconstruction of surface pCO_2 in the ideal ('bgcArgo') sampling scheme, a mean and growing offset in CO_2 flux of 0.4 to 0.6 PgC yr^{-1} at the end of the time-series remains between the 'known-truth' and the products when using the products native gas-transfer velocity. For MPI-SOM-FFN, this offset can be attributed to methodological biases and gas-exchange uncertainty in about equal amounts. For CarboScope, the discrepancy is entirely due to choices in gas-exchange calculation. All contributions work towards a discrepancy particularly in the high latitudes. This is in agreement with reported uncertainties of the pCO_2 observation-based estimates of 0.45 PgC yr^{-1}–0.6 PgC yr^{-1} with a dominant contribution from the gas-transfer velocity [48,53,55,58].

The 1-σ uncertainty of pCO_2-based CO_2 flux estimates in the Global Carbon Budget [2] is quantified as a contribution of random uncertainty (0.3 PgC yr^{-1}, standard deviation of pCO_2 product ensemble) uncertainty in pCO_2 measurements (0.2 PgC yr^{-1}, [59]), gas-transfer velocity k (0.2 PgC yr^{-1}, [59], wind product (0.1 PgC yr^{-1}, [55]), uncertainty in river flux adjustment (0.3 PgC yr^{-1}, 2-σ [5]) and pCO_2 mapping (0.2 PgC yr^{-1}, [23]), resulting in a total 1-σ uncertainty of 0.6 PgC yr^{-1}. Our analysis illustrates that the total uncertainty is increasing over time and may be 0.6 PgC yr^{-1} even when the river flux adjustment is not needed as in our study design. Based on the subsampling experiments conducted here, the mapping uncertainty may be an underestimate and an additional uncertainty term due to sparse and skewed data distribution of 0.2 PgC yr^{-1} (difference between mean flux in SOCAT and ideal sampling schemes) should be added to this assessment (§3a, figure 5).

The discrepancy between the ocean carbon sink estimates by the ten global ocean biogeochemistry models (such as FESOM-REcoM) and the 8 surface pCO_2-based data products (such as MPI-SOM-FFN and CarboScope) in the Global Carbon Budget is large [2]. While they matched reasonably well in the 1990s, they have diverged after 2000. The full range of all estimates amounts to 2 PgC yr^{-1} in 2021, and the difference between the GOBM and pCO_2-product ensemble means to 1 PgC yr^{-1}. It is estimated that the GOBMs underestimate the ocean carbon sink by 10–20% as evidenced by a direct comparison to interior ocean anthropogenic carbon accumulation [2,3]. This is also supported by higher estimates from atmospheric inversions (although they often use pCO_2 products as priors, and are thus equally affected by the uncertainties studied here) and $O_2 : N_2$ ratios [2,4]. Analysis of Earth System Models (which may differ from the biases in GOBMs forced by atmospheric reanalysis) indicate that this underestimation may be explained by biases in ocean carbon transport and mixing [11–13], biases in the chemical buffer capacity [15], and by a late starting date of the simulations [60]. Here, we demonstrate that the divergence since the early 2000s can to a large degree be explained by the overestimation of the trend by the pCO_2 products. This effect is stronger in the neural network method than in CarboScope. The effect of sparse and unevenly distributed observations on other neural network and cluster regression methods that contribute to the Global Carbon Budget [24,28,61] remains to be tested. The MPI-SOM-FFN product shows by far the largest CO_2 flux trend since 2000 compared with the other GCB2022 pCO_2 products (table 1), although the exact numbers depend on chosen start date. However, cluster-regression (OS-ETHZ-GRaCER) and neural network (CMEMS-LESCE-FFNN) approaches can also have trend estimates on the lower end of the spectrum (table 1), reflecting the role of methodological choices within these methods. Yet, all pCO_2 products show a larger trend than the GOBM ensemble mean and also higher than the GOBM with the largest trend (table 1). Notably, the standard deviation of the *pCO_2* products' trends is an order of magnitude larger than of the GOBMs, illustrating that the trend is poorly constrained by the pCO_2 products. The trend in the GOBM FESOM-1.4-REcoM (0.43 PgC yr^{-1} decade^{-1}) that is used here, falls close to the GOBM ensemble average (0.41 PgC yr^{-1} decade^{-1}) and close to the successor version FESOM-2.1-REcoM2 (0.45 PgC yr^{-1} decade^{-1}). This indicates that the data distribution plays a role in the full set of observational products within the Global Carbon Budget. A pCO_2-product based on pCO_2 mismatches between GOBMs and observations [62] clusters with a few products estimating a somewhat slower growing ocean carbon sink, but still close to the product ensemble average and well above any GOBM estimate (table 1). In conclusion, we advocate for continued

pCO_2 mapping intercomparisons [29] and observation-system design experiments [31,32] to understand the biases in all mapping methods. The ensemble of global ocean biogeochemistry models forced with atmospheric reanalysis and observed atmospheric CO_2 as used in the GCB are closer to interannual climate and carbon variability than Earth System Models, and thus are an ideal tool for such experiments. While the results on data distribution may be dependent on the simulated pCO_2 in the ocean model used, the subsampling experiment demonstrates the sensitivity of the mapping products to sparse and unevenly distributed observations, as well as uncertainties in gas-exchange choices.

In terms of decadal and interannual variability, the pCO_2 products are skilful when driven with sufficient and evenly distributed pCO_2 observations. We note that the metric for amplitude of variability (standard deviation of detrended time-series) does not separate between interannual and decadal variability and thus needs to be interpreted with some caution. Inspection of the time-series (figure 4) indicates the reduction of erroneous decadal variability and slight improvement in interannual variability in the ideal bgcArgo sampling scheme (except for MPI-SOM-FFN in the tropics). We further note that also the estimate of variability and trend is affected by choices in gas-exchange calculations.

In agreement with previous studies of observation system design [31–33], we emphasize the need for a sustained and coherent network of high-quality pCO_2 observations. Ship-based measurements provide high-accuracy data, but are too sparse in many regions of the ocean, especially in the Southern Hemisphere. A combination of multiple platforms will be needed including autonomous devices, such as biogeochemical Argo floats, but also saildrones [37], wave gliders, moorings. We demonstrate that an observation network with a coverage similar to the planned global deployment of 1000 biogeochemical Argo floats may be sufficient to fill the gaps in the current ship-based networks. In reality, the network will not follow such a regular distribution, and high accuracy pCO_2 measurements will be needed to at least partially fill the gaps in the Southern Ocean. The bgcArgo floats are equipped with pH sensors, and hence pCO_2 needs to be derived from pH and a multi-linear regression-derived alkalinity estimate. Currently, this procedure is associated with uncertainties that are substantially larger than for the pCO_2 measurements on ships [18,35,63]. Our analysis reveals that systematic biases of around $5\,\mu atm$ can lead to large errors in reconstructed pCO_2 and CO_2 flux. As the array of biogeochemical Argo floats expands, it will be essential to conduct on-going intercomparisons between float and ship-based measurements with the goal of developing an unbiased multi-platform sampling array that captures the relevant spatial and temporal variability. This may involve more extensive sampling of surface pH and alkalinity on research vessels and ships of opportunity to allow better assessment of float pH and salinity-derived alkalinity and targeting float locations for purposeful crossover comparisons.

5. Conclusion

We have assessed the sensitivity of surface ocean pCO_2 reconstructions to sparse observations and uneven sampling in space and time. The Southern Ocean decadal variability stands out as being especially prone to such uncertainties, and this is stronger in the neural network compared with the process-based mapping scheme. We thus caution the reader to not relate to pCO_2 products as observations, but to acknowledge that they are statistical models of sparse observations. There are two possible ways forward that both need to be addressed in order to improve robustness of the ocean carbon monitoring system. First, the observational network must be strengthened, with sustained and operational funding and with a combination of multiple platforms. Notably, the total number of global high-quality ocean CO_2 observations has surpassed the number of observations in our tested 'ideal' sampling design, indicating that the considerate distribution of sampling sites is more important than simply adding more observations in the same places. Secondly, the mapping methods should be routinely evaluated for their ability to handle skewed data distribution, and should be improved when necessary.

References

1. Hauck J *et al.* 2020 Consistency and challenges in the ocean carbon sink estimate for the global carbon budget. *Front. Mar. Sci.* **7**, 571720. (doi:10.3389/fmars.2020.571720)

2. Friedlingstein P *et al.* 2022 Global carbon budget 2021. *Earth Syst. Sci. Data* **14**, 1917–2005. (doi:10.5194/essd-14-1917-2022)

3. Gruber N *et al.* 2019 The oceanic sink for anthropogenic CO_2 from 1994 to 2007. *Science* **363**, 1193–1199. (doi:10.1126/science.aau5153)

4. Tohjima Y, Mukai H, Machida T, Hoshina Y, Nakaoka SI. 2019 Global carbon budgets estimated from atmospheric O_2 to N_2 and CO_2 observations in the western Pacific region over a 15year period. *Atmos. Chem. Phys.* **19**, 9269–9285. (doi:10.5194/acp-19-9269-2019)

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

6. Aumont O, Orr JC, Monfray P, Ludwig W, Amiotte-Suchet P, Probst JL. 2001 Riverine-driven interhemispheric transport of carbon. *Global Biogeochem. Cycles* **15**, 393–405. (doi:10.1029/1999GB001238)

7. 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**, 55–88. (doi:10.5194/bg-17-55-2020)

8. Fennel K, Mattern JP, Doney SC, Bopp L, Moore AM, Wang B, Yu L. 2022 Ocean biogeochemical modelling. *Nat. Rev. Methods Primers* **2**, 76. (doi:10.1038/s43586-022-00154-2)

9. Tsujino H *et al.* 2018 JRA-55 based surface dataset for driving ocean–sea-ice models (JRA55-do). *Ocean Modell.* **130**, 79–139. (doi:10.1016/j.ocemod.2018.07.002)

10. Denman KL *et al.* 2007 Couplings between changes in the climate system and biogeochemistry. In *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change* (eds S Solomon, D Qin, M Manning, M Marquis, K Averyt, MMB Tignor, HL Miller, ZL Chen), pp. 499–587. Cambridge, UK: Cambridge University Press.

11. Goris N, Tjiputra JF, Olsen A, Schwinger J, Lauvset SK, Jeansson E. 2018 Constraining projection-based estimates of the future North Atlantic carbon uptake. *J. Climate* **31**, 3959–3978. (doi:10.1175/JCLI-D-17-0564.1)

12. Terhaar J, Frölicher TL, Joos F. 2021 Southern Ocean anthropogenic carbon sink constrained by sea surface salinity. *Sci. Adv.* **7**, eabd5964. (doi:10.1126/sciadv.abd5964)

13. Bourgeois T, Goris N, Schwinger J, Tjiputra JF. 2022 Stratification constrains future heat and carbon uptake in the Southern Ocean between 30S and 55S. *Nat. Commun.* **13**, 340. (doi:10.1038/s41467-022-27979-5)

14. Vaittinada Ayar P *et al.* 2022 Contrasting projections of the ENSO-driven CO_2 flux variability in the equatorial Pacific under high-warming scenario. *Earth Syst. Dyn.* **13**, 1097–1118. (doi:10.5194/esd-13-1097-2022)

15. Terhaar J, Frölicher TL, Joos F. 2022 Observation-constrained estimates of the global ocean carbon sink from Earth system models. *Biogeosciences* **19**, 4431–4457. (doi:10.5194/bg-19-4431-2022)

16. Pfeil B *et al.* 2013 A uniform, quality controlled surface ocean CO_2 atlas (SOCAT). *Earth Syst. Sci. Data* **5**, 125–143. (doi:10.5194/essd-5-125-2013)

17. Sabine CL *et al.* 2013 Surface ocean CO_2 Atlas (SOCAT) gridded data products. *Earth Syst. Sci. Data* **5**, 145–153. (doi:10.5194/essd-5-145-2013)

18. Bakker DCE *et al.* 2016 A multi-decade record of high-quality CO_2 data in version 3 of the Surface Ocean CO_2 Atlas (SOCAT). *Earth Syst. Sci. Data* **8**, 383–413. (doi:10.5194/essd-8-383-2016)

19. Ho DT, Law CS, Smith MJ, Schlosser P, Harvey M, Hill P. 2006 Measurements of air-sea gas exchange at high wind speeds in the Southern Ocean: implications for global parameterizations. *Geophys. Res. Lett.* **33**, L16611. (doi:10.1029/2006GL026817)

20. Wanninkhof R. 2014 Relationship between wind speed and gas exchange over the ocean revisited: gas exchange and wind speed over the ocean. *Limnol. Oceanogr.: Methods* **12**, 351–362. (doi:10.4319/lom.2014.12.351)

21. Rödenbeck C, Keeling RF, Bakker DCE, 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 Sci.* **9**, 193–216.

22. Landschützer P, Gruber N, Bakker DCE, Schuster U, Nakaoka S, Payne MR, Sasse TP, Zeng J. 2013 A neural network-based estimate of the seasonal to inter-annual variability of the Atlantic Ocean carbon sink. *Biogeosciences* **10**, 7793–7815.

23. Landschützer P, Gruber N, Bakker DCE, Schuster U. 2014 Recent variability of the global ocean carbon sink. *Global Biogeochem. Cycles* **28**, 927–949.

24. Zeng J, Nojiri Y, Landschützer P, Telszewski M, Nakaoka S. 2014 A global surface ocean fCO_2 climatology based on a feed-forward neural network. *J. Atmos. Oceanic Technol.* **31**, 1838–1849. (doi:10.1175/JTECH-D-13-00137.1)

25. Denvil-Sommer A, Gehlen M, Vrac M, Mejia C. 2019 LSCE-FFNN-v1: a two-step neural network model for the reconstruction of surface ocean pCo_2 over the global ocean. *Geosci. Model Dev.* **12**, 2091–2105. (doi:10.5194/gmd-12-2091-2019)

26. Gregor L, Lebehot AD, Kok S, Scheel Monteiro PM. 2019 A comparative assessment of the uncertainties of global surface ocean CO2 estimates using a machine-learning ensemble (CSIR-ML6 version 2019a) – have we hit the wall?. *Geosci. Model Dev.* **12**, 5113–5136. (doi:10.5194/gmd-12-5113-2019)

27. 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. *J. Oceanogr.* **77**, 323–358. (doi:10.1007/s10872-020-00571-5)

28. 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 Syst. Sci. Data* **13**, 777–808. (doi:10.5194/essd-13-777-2021)

29. Rödenbeck C *et al.* 2015 Data-based estimates of the ocean carbon sink variability–first results of the Surface Ocean pCO_2 Mapping intercomparison (SOCOM). *Biogeosciences* **12**, 7251–7278.

30. Bushinsky SM, Landschützer P, Rödenbeck C, Gray AR, Baker D, Mazloff MR, Resplandy L, Johnson KS, Sarmiento JL. 2019 Reassessing southern ocean air–sea CO_2 flux estimates with the addition of biogeochemical float observations. *Global Biogeochem. Cycles* **33**, 1370–1388. (doi:10.1029/2019GB006176)

31. Gloege L *et al.* 2021 Quantifying errors in observationally based estimates of ocean carbon sink variability. *Global Biogeochem. Cycles* **35**, e2020GB006788. (doi:10.1029/2020GB006788)

32. Denvil-Sommer A, Gehlen M, Vrac M. 2021 Observation system simulation experiments in the Atlantic Ocean for enhanced surface ocean co2 reconstructions. *Ocean Sci.* **17**, 1011–1030. (doi:10.5194/os-17-1011-2021)

33. Djeutchouang LM, Chang N, Gregor L, Vichi M, Monteiro PMS. 2022 The sensitivity of pCO$_2$ reconstructions to sampling scales across a Southern Ocean sub-domain: a semi-idealized ocean sampling simulation approach. *Biogeosciences* **19**, 4171–4195. (doi:10.5194/bg-19-4171-2022)

34. Le Quéré C *et al.* 2014 Global carbon budget 2013. *Earth Syst. Sci. Data* **6**, 235–263.

35. Williams NL *et al.* 2017 Calculating surface ocean pCO$_2$ from biogeochemical Argo floats equipped with pH: an uncertainty analysis: calculating ocean pCO$_2$ From Float pH. *Global Biogeochem. Cycles* **31**, 591–604. (doi:10.1002/2016GB005541)

36. Roemmich D *et al.* 2019 On the future of argo: a global, full-depth, multi-disciplinary array. *Front. Mar. Sci.* **6**, 439. (doi:10.3389/fmars.2019.00439)

37. Sutton AJ, Williams NL, Tilbrook B. 2021 Constraining southern ocean CO$_2$ flux uncertainty using uncrewed surface vehicle observations. *Geophys. Res. Lett.* **48**, 421. (doi:10.1029/2020GL091748)

38. Gray AR, Johnson KS, Bushinsky SM, Riser SC, Russell JL, Talley LD, Wanninkhof R, Williams NL, Sarmiento JL. 2018 Autonomous biogeochemical floats detect significant carbon dioxide outgassing in the high–latitude southern ocean. *Geophys. Res. Lett.* **45**, 9049–9057. (doi:10.1029/2018GL078013)

39. Wang Q, Danilov S, Sidorenko D, Timmermann R, Wekerle C, Wang X, Jung T, Schröter J. 2014 The finite element sea ice-ocean model (FESOM) v.1.4: formulation of an ocean general circulation model. *Geosci. Model Dev.* **7**, 663–693. (doi:10.5194/gmd-7-663-2014)

40. Hauck J, Völker C, Wang T, Hoppema M, Losch M, Wolf-Gladrow DA. 2013 Seasonally different carbon flux changes in the Southern Ocean in response to the southern annular mode. *Global Biogeochem. Cycles* **27**, 1236–1245. (doi:10.1002/2013GB004600)

41. Schourup-Kristensen V, Sidorenko D, Wolf-Gladrow DA, Völker C. 2014 A skill assessment of the biogeochemical model REcoM2 coupled to the finite element sea ice–ocean model (FESOM 1.3). *Geosci. Model Dev.* **7**, 2769–2802. (doi:10.5194/gmd-7-2769-2014)

42. Garcia HE *et al.* 2013 World ocean atlas 2013. volume 4, dissolved inorganic nutrients (phosphate, nitrate, silicate). *NOAA atlas NESDIS*. (doi:10.7289/V5J67DWD)

43. Lauvset SK *et al.* 2016 A new global interior ocean mapped climatology: the 1° × 1° GLODAP version 2. *Earth Syst. Sci. Data* **8**, 325–340. (doi:10.5194/essd-8-325-2016)

44. Friedlingstein P *et al.* 2020 Global carbon budget 2020. *Earth Syst. Sci. Data* **12**, 3269–3340. (doi:10.5194/essd-12-3269-2020)

45. Dlugokencky E, Tans P. 2019 Trends in atmospheric carbon dioxide. *National Oceanic & Atmospheric Administration, Earth System Research Laboratory (NOAA/ESRL)*. See https://gml.noaa.gov/ccgg/trends/global.html. Last access: 25 September 2022.

46. Ballantyne A, Alden C, Miller J, Tans P, White J. 2012 Increase in observed net carbon dioxide uptake by land and oceans during the past 50 years. *Nature* **488**, 70–72. (doi:10.1038/nature11299)

47. Orr JC, Epitalon JM. 2015 Improved routines to model the ocean carbonate system: mocsy 2.0. *Geosci. Model Dev.* **8**, 485–499. (doi:10.5194/gmd-8-485-2015)

48. Rödenbeck C, DeVries T, Hauck J, Le Quéré C, Keeling RF. 2022 Data-based estimates of interannual sea–air CO$_2$ flux variations 1957–2020 and their relation to environmental drivers. *Biogeosciences* **19**, 2627–2652.

49. Landschützer P, Gruber N, Bakker DCE. 2016 Decadal variations and trends of the global ocean carbon sink. *Global Biogeochem. Cycles* **30**, 1396–1417.

50. Lee K *et al.* 2006 Global relationships of total alkalinity with salinity and temperature in surface waters of the world's oceans. *Geophys. Res. Lett.* **33**, L19605. (doi:10.1029/2006GL027207)

51. DeVries T. 2022 Atmospheric CO$_2$ and sea surface temperature variability cannot explain recent Decadal variability of the ocean CO$_2$ Sink. *Geophys. Res. Lett.* **49**, 21084–21112. (doi:10.1029/2021GL096018)

52. Wanninkhof R. 1992 Relationship between wind speed and gas exchange over the ocean. *J. Geophys. Res.* **97**, 7373. (doi:10.1029/92JC00188)

53. Naegler T. 2009 Reconciliation of excess ^{14}C-constrained global CO$_2$ piston velocity estimates. *Tellus B: Chem. Phys. Meteorol.* **61**, 372–384. (doi:10.1111/j.1600-0889.2008.00408.x)

54. Weiss R. 1974 Carbon dioxide in water and seawater: the solubility of a non-ideal gas. *Mar. Chem.* **2**, 203–215. (doi:10.1016/0304-4203(74)90015-2)

55. Fay AR *et al.* 2021 SeaFlux: harmonization of air–sea CO_2 fluxes from surface pCO_2 data products using a standardized approach. *Earth Syst. Sci. Data* **13**, 4693–4710. (doi:10.5194/essd-13-4693-2021)

56. Sarmiento JL, Orr JC, Siegenthaler U. 1992 A perturbation simulation of CO_2 uptake in an ocean general circulation model. *J. Geophys. Res.* **97**, 3621. (doi:10.1029/91JC02849)

57. Landschützer P *et al.* 2015 The reinvigoration of the Southern Ocean carbon sink. *Science* **349**, 1221–1224.

58. Woolf D *et al.* 2019 Key Uncertainties in the recent air–sea flux of CO_2. *Global Biogeochem. Cycles* **33**, 1548–1563. (doi:10.1029/2018GB006041)

59. Wanninkhof R *et al.* 2013 Global ocean carbon uptake: magnitude, variability and trends. *Biogeosciences* **10**, 1983–2000. (doi:10.5194/bg-10-1983-2013)

60. Bronselaer B, Winton M, Russell J, Sabine CL, Khatiwala S. 2017 Agreement of CMIP5 simulated and observed ocean anthropogenic CO_2 uptake. *Geophys. Res. Lett.* **44**, 12–298. (doi:10.1002/2017GL074435)

61. Chau TTT, 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**, 1087–1109. (doi:10.5194/bg-19-1087-2022)

62. Bennington V, Gloege L, McKinley GA. 2022 Variability in the global ocean carbon sink From 1959 to 2020 by correcting models with observations. *Geophys. Res. Lett.* **49**, e2022GL098632. (doi:10.1029/2022GL098632)

63. Williams NL, Juranek LW, Johnson KS, Feely RA, Riser SC, Talley LD, Russell JL, Sarmiento JL, Wanninkhof R. 2016 Empirical algorithms to estimate water column pH in the Southern Ocean. *Geophys. Res. Lett.* **43**, 3415–3422. (doi:10.1002/2016GL068539)

64. Hauck J, Nissen C, Landschützer P, Rödenbeck C, Bushinsky S, Olsen A. 2023 Sparse observations induce large biases in estimates of the global ocean CO_2 sink: an ocean model subsampling experiment. Zenodo. (doi:10.5281/zenodo.7784745)

65. Hauck J, Nissen C, Landschützer P, Rödenbeck C, Bushinsky S, Olsen A. 2023 Sparse observations induce large biases in estimates of the global ocean CO_2 sink: an ocean model subsampling experiment. Supplementary figures. Figshare. (doi:10.6084/m9.figshare.c.6502553)