

# Geophysical Research Letters®

## RESEARCH LETTER

10.1029/2024GL108502

## Accuracy of Ocean CO<sub>2</sub> Uptake Estimates at a Risk by a Reduction in the Data Collection



### Key Points:

- Lower surface ocean  $f\text{CO}_2$  data availability leads to higher uncertainty in data-based estimates of ocean CO<sub>2</sub> uptake
- The long-term trend in the ocean CO<sub>2</sub> flux increases by 1.5 times for subsequent years if the data availability is reduced to that in 2000
- The annual mean CO<sub>2</sub> flux is not sensitive to the seasonal skew in the data and to the addition of low accuracy data

### Supporting Information:

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

### Correspondence to:

Y. Dong,  
[ydong@geomar.de](mailto:ydong@geomar.de)

### Citation:

Dong, Y., Bakker, D. C. E., & Landschützer, P. (2024). Accuracy of ocean CO<sub>2</sub> uptake estimates at a risk by a reduction in the data collection. *Geophysical Research Letters*, 51, e2024GL108502. <https://doi.org/10.1029/2024GL108502>

Received 30 JAN 2024

Accepted 13 APR 2024

### Author Contributions:

**Conceptualization:** Yuanxu Dong, Dorothee C. E. Bakker  
**Data curation:** Dorothee C. E. Bakker  
**Formal analysis:** Yuanxu Dong  
**Funding acquisition:** Dorothee C. E. Bakker  
**Investigation:** Yuanxu Dong  
**Methodology:** Yuanxu Dong, Dorothee C. E. Bakker, Peter Landschützer  
**Project administration:** Yuanxu Dong  
**Resources:** Dorothee C. E. Bakker, Peter Landschützer  
**Software:** Peter Landschützer  
**Supervision:** Dorothee C. E. Bakker  
**Validation:** Dorothee C. E. Bakker, Peter Landschützer  
**Visualization:** Yuanxu Dong

Yuanxu Dong<sup>1,2</sup> , Dorothee C. E. Bakker<sup>1</sup> , and Peter Landschützer<sup>3</sup> 

<sup>1</sup>Centre for Ocean and Atmospheric Sciences, School of Environmental Sciences, University of East Anglia, Norwich, UK, <sup>2</sup>GEOMAR Helmholtz Centre for Ocean Research Kiel, Kiel, Germany, <sup>3</sup>Flanders Marine Institute (VLIZ), Ostend, Belgium

**Abstract** Observation-based quantification of ocean carbon dioxide (CO<sub>2</sub>) uptake relies on synthesis data sets such as the Surface Ocean CO<sub>2</sub> Atlas (SOCAT). However, the data collection effort has dramatically declined and the number of annual data sets in SOCATv2023 decreased by ~35% from 2017 to 2021. This decline has led to a 65% increase (from 0.15 to 0.25 Pg C yr<sup>-1</sup>) in the standard deviation of seven SOCAT-based air-sea CO<sub>2</sub> flux estimates. Reducing the availability of the annual data to that in the year 2000 creates substantial bias (50%) in the long-term flux trend. The annual mean CO<sub>2</sub> flux is insensitive to the seasonal skew of the SOCAT data and to the addition of the lower accuracy data set available in SOCAT. Our study highlights the need for sustained data collection and synthesis, to inform the Global Carbon Budget assessment, the UN-led climate negotiations, and measurement, reporting, and verification of ocean-based CO<sub>2</sub> removal projects.

**Plain Language Summary** The Surface Ocean CO<sub>2</sub> Atlas (SOCAT) data set plays a crucial role in estimating the ocean carbon sink component of the Global Carbon Budget. However, the number of data sets available in SOCAT each year has drastically decreased since 2017. This study shows that the uncertainty in the data-based ocean CO<sub>2</sub> flux estimate has increased by 65% due to this decline in data availability. The estimated fluxes, especially the long-term flux trend, are remarkably affected by the data availability in SOCAT, reducing the reliability of ocean CO<sub>2</sub> uptake estimates in years and regions with sparse observations.

## 1. Introduction

The carbon dioxide (CO<sub>2</sub>) emitted by human activities is the major contributor to climate change (IPCC, 2021). The ocean is crucial in slowing down global warming by taking up a quarter of the anthropogenic CO<sub>2</sub> emissions (Friedlingstein et al., 2023). Quantifying the air-sea CO<sub>2</sub> flux and understanding its variability is critical for predicting the future climate and the ocean environment and developing climate mitigation strategies (Lee et al., 2023).

Previous studies about global ocean CO<sub>2</sub> uptake mainly relied on models (Le Quéré et al., 2014). Since 2011, the public release of the updates to the Surface Ocean CO<sub>2</sub> Atlas (SOCAT, Bakker et al., 2016; Pfeil et al., 2013) has significantly advanced our ability to quantify the ocean CO<sub>2</sub> sink based on observations. The latest SOCAT release, SOCATv2023, provides 42.8 million quality-controlled, in situ surface ocean CO<sub>2</sub> fugacity ( $f\text{CO}_2$ ) measurements over the period 1957 through 2022 (Bakker, Alin, et al., 2023). The SOCAT data are primarily shipboard measurements typically at a 1-min sampling frequency. High-quality  $f\text{CO}_2$  observations (35.6 million, estimated uncertainty <5  $\mu\text{atm}$ ) are included in the synthesis files and used for producing a 1° by 1°, monthly gridded product (Sabine et al., 2013). A variety of mapping methods have subsequently been developed and used to interpolate the sparse gridded  $f\text{CO}_2$  data to yield gap-free, monthly  $f\text{CO}_2$ -products for the global ocean (e.g., Landschützer et al., 2013; Rödenbeck et al., 2013). These  $f\text{CO}_2$  products combined with the atmospheric CO<sub>2</sub> mole fraction (e.g., Dlugokencky & Tans, 2023) and a gas transfer velocity parametrization (e.g., Wanninkhof, 2014) produce observation-based estimates of contemporary ocean CO<sub>2</sub> uptake.

The SOCAT data thus enable novel estimates of the global air-sea CO<sub>2</sub> flux at a monthly timescale, which enables to address the decadal, interannual, and seasonal variability of the ocean carbon sink (DeVries et al., 2023; Gruber et al., 2023; Rodgers et al., 2023). The SOCAT products have been used to report the annual CO<sub>2</sub> sink of the ocean component in the Global Carbon Budget (Friedlingstein et al., 2023) and have indicated a stagnation of the Southern Ocean CO<sub>2</sub> sink from 1980s to the early 2000s and identified subsequent strong reinvigoration (Landschützer et al., 2015), although with disagreement regarding the strength of the uptake in the post 2000

© 2024. The Authors.

This is an open access article under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

Writing – original draft: Yuanxu Dong  
Writing – review & editing: Dorothee  
C. E. Bakker, Peter Landschützer

period (DeVries et al., 2019; Ritter et al., 2017). In addition, the SOCAT data provide valuable information for determining the seasonal variability of the CO<sub>2</sub> sink in different ocean basins and identifying the difficulty of models in capturing biological processes (Hauck, Nissen, et al., 2023; Rodgers et al., 2023) that can act on multiple timescales (Ostle et al., 2022). Furthermore, the SOCAT-based assessment of the variability in the air-sea CO<sub>2</sub> flux has advanced our understanding of the natural component of ocean CO<sub>2</sub> uptake (Gruber et al., 2023).

Although the SOCAT synthesis plays a critical role in examining the global ocean carbon cycle, the *f*CO<sub>2</sub> data collection effort it relies on is at risk, as the number of monthly data sets and thus the number of gridded *f*CO<sub>2</sub> values has sharply decreased since 2017 (Bakker, Alin, et al., 2023). The impact of this data decline on estimates of ocean CO<sub>2</sub> uptake is not clear. In addition, the SOCAT data are very sparse (especially before 2000) and strongly skewed to the summer season (especially in the Southern Ocean), which may create bias in the seasonal or longer time scale flux estimates (Denvil-Sommer et al., 2021; DeVries et al., 2023; Djetchouang et al., 2022; Gloege et al., 2021; Hauck, Nissen, et al., 2023). Furthermore, only *f*CO<sub>2</sub> data in SOCAT with an estimated accuracy better than 5 μatm (data set flags of A, B, C, and D) are routinely used to estimate the CO<sub>2</sub> flux, while the *f*CO<sub>2</sub> data with a flag of E (estimated uncertainty between 5 and 10 μatm, 7.2 million values) are generally not used for the flux estimates (Bakker et al., 2016). In the context of the decrease in the SOCAT synthesis data, the increasing number of flag E data might provide a valuable source of additional *f*CO<sub>2</sub> data in flux estimates.

In this study, we design experiments to assess the impact of the reduction in the availability of *f*CO<sub>2</sub> data sets, the seasonal skew of the SOCAT data, and the addition of the SOCAT flag E data on the air-sea CO<sub>2</sub> flux estimates.

## 2. Experimental Setup

The SOCAT-based *f*CO<sub>2</sub>-products estimate the air-sea CO<sub>2</sub> flux with the bulk equation:

$$F = K_{660}(Sc/660)^{-0.5}(\alpha_w fCO_{2w} - \alpha_i fCO_{2a}) \quad (1)$$

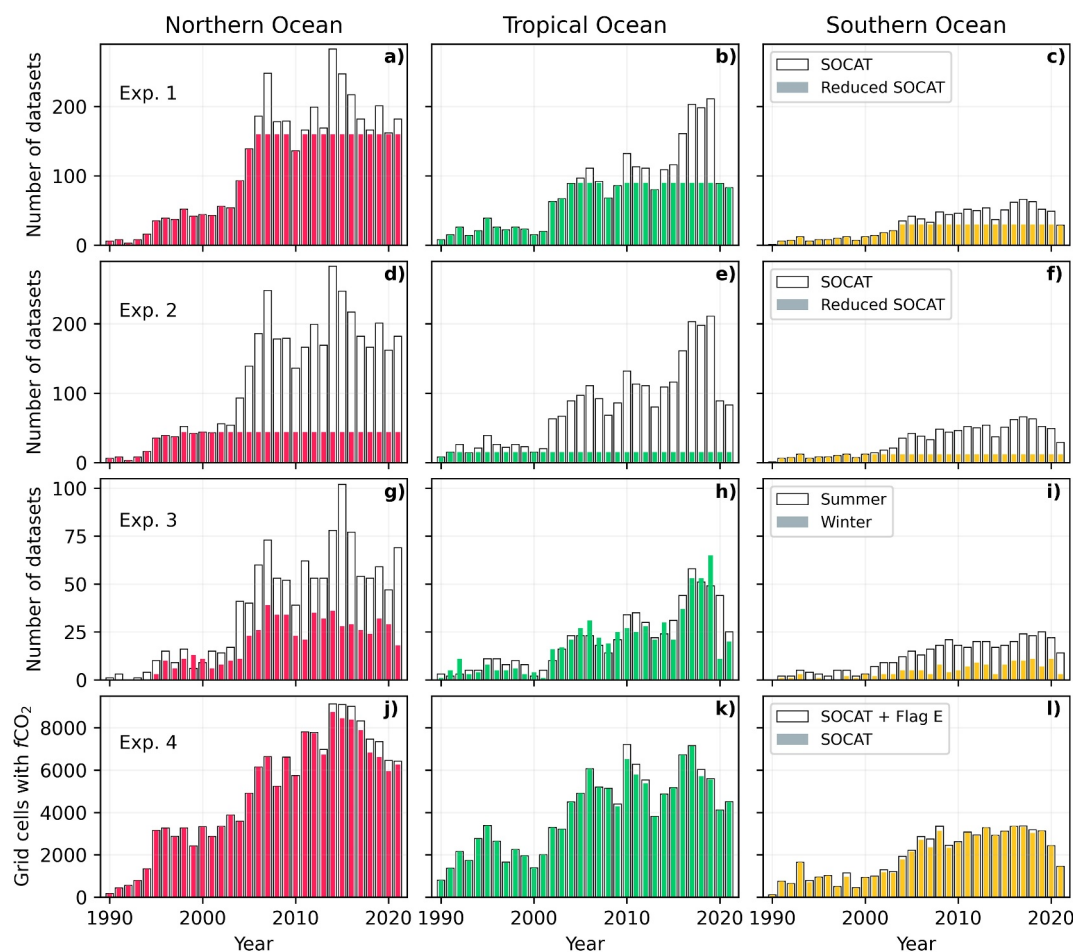
where *F* (mmol m<sup>-2</sup> day<sup>-1</sup>) is the air-sea CO<sub>2</sub> flux and *Sc* is the non-dimensional Schmidt number. *Sc* is equal to 660 for CO<sub>2</sub> at 20°C and 35 psu seawater. *K*<sub>660</sub> (cm h<sup>-1</sup>) is the gas transfer velocity (e.g., Wanninkhof, 2014) normalized to a *Sc* of 660. The CO<sub>2</sub> solubility (mol L<sup>-1</sup> atm<sup>-1</sup>) at the subsurface and the air-sea interface are represented by α<sub>w</sub> and α<sub>i</sub>, respectively (Dong et al., 2022; Watson et al., 2020; Woolf et al., 2016). *Sc* and α are calculated from seawater temperature and salinity (Wanninkhof, 2014; Weiss, 1974). The CO<sub>2</sub> fugacity (μatm) at the subsurface and just above the air-sea interface are represented by *f*CO<sub>2w</sub> and *f*CO<sub>2a</sub>, respectively. To balance the dimensions of the left and the right side for the given units, a factor of 0.24 should be multiplied to the right side of Equation 1.

In four experiments, we use a neural network-based interpolation method (Landschützer et al., 2013) to map the gridded or re-gridded *f*CO<sub>2w</sub> data in SOCATv2023 (see Experiments 1–4 below) to yield 1° by 1°, monthly gap-free *f*CO<sub>2w</sub> products. A summary of the input data for neural network training and the gas transfer velocity is in Text S1 of the Supporting Information S1. The latest SOCAT version 2023 (SOCAT hereafter) is used for the four experiments. For convenience of the analysis, we divide the global ocean into three ocean regions: the Northern Ocean (defined as north of 30°N), the Tropical Ocean (defined as 30°S–30°N), and the Southern Ocean (defined as south of 30°S). We divide all the data in SOCAT into the Northern, the Tropical, and the Southern Oceans according to their latitude.

### Experiment 1: Sensitivity of the flux estimate to the recent reduction in data availability

SOCATv2023 highlights a striking decline in the number of 1° by 1°, monthly grid cells with *f*CO<sub>2</sub> since 2017 (Bakker, Alin, et al., 2023). This decrease started 2 years before the pandemic (i.e., 2020), with the pandemic exacerbating the decrease. The decrease in the number of monthly grid cells with *f*CO<sub>2</sub> reflects a decrease in the number of data sets collected in the open ocean in recent years (Figure 1).

Considering that the number of data sets in the final year (i.e., 2022 in SOCATv2023) is always relatively low compared to the previous years due to delays in data submission, we take the number of data sets in the year 2021 in SOCAT as a reference. We then randomly reduce the number of data sets in the Northern, Tropical, and Southern Oceans in years with more than that in 2021, respectively (Figures 1a–1c). In total, we remove 565 (17%), 574 (28%), and 321 (38%) data sets from the Northern, Tropical, and Southern Oceans, respectively from



**Figure 1.** The number of data sets or grid cells with  $f\text{CO}_2$  in SOCATv2023 for each year from 1990 to 2021. The left, middle, and right panels represent the Northern, Tropical, and Southern Oceans, respectively. (a–c) White bars represent the number of data sets in Surface Ocean  $\text{CO}_2$  Atlas (SOCAT), and shaded bars denote the number of remaining data sets after the removal of some data sets to match a similar number in 2021 (see Section 2, experiment 1). (d–f) White bars represent the number of data sets in SOCAT while shaded bars correspond to the number of the remaining data sets after removal of data sets more than the number of those in the year 2000 (see Section 2, experiment 2). (g–i) White and shaded bars indicate the number of data sets in summer and winter, respectively (see Section 2, experiment 3). (j–l) Solid bars represent the number of monthly grid cells with  $f\text{CO}_2$  with an estimated accuracy better than  $5 \mu\text{atm}$  in SOCAT, while white bars indicate the number of grid cells upon addition of lower accuracy data sets with a flag of E (see Section 2, experiment 4). The white bars in all subplots start at the bottom (i.e., zero).

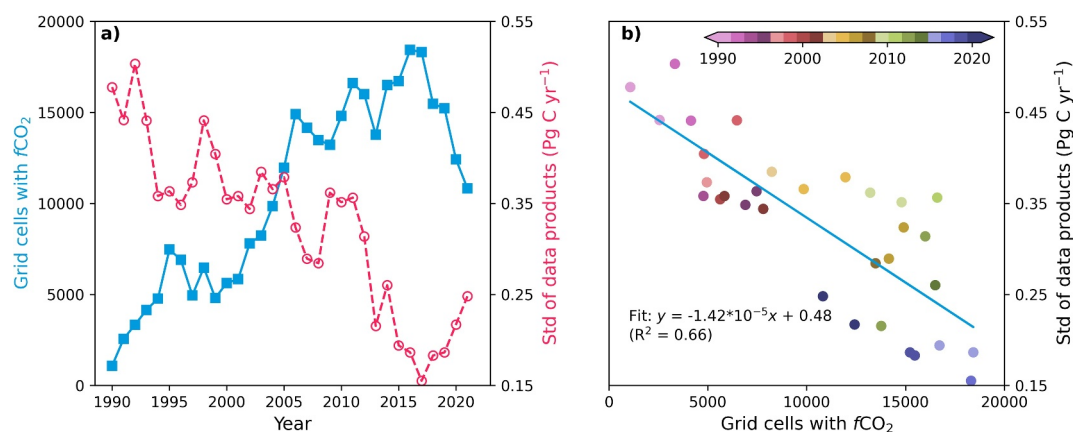
2005 to 2021. We then re-grid the reduced SOCAT synthesis data set to yield an additional  $1^\circ$  by  $1^\circ$ , monthly data set, which is used for the interpolation and the flux calculation.

### Experiment 2: Sensitivity of the flux trend to data availability

The long-term (decadal) trends of the air-sea  $\text{CO}_2$  flux are not well-constrained (Friedlingstein et al., 2023). To test if data availability affects the trend in the SOCAT-based flux estimates, we take the number of data sets in 2000 of SOCAT as a reference, and randomly reduce the data sets in years with more than the number of data sets in 2000 in the Northern, Tropical, and Southern Oceans, respectively (Figures 1d–1f). We then re-grid the reduced SOCAT synthesis data set and interpolate it for the flux calculation.

### Experiment 3: Sensitivity of the flux estimate to seasonal skew in the data

The  $f\text{CO}_2$  observations in SOCAT are seasonally skewed with more data in summer and less data in winter in the high latitude regions (Figures 1g and 1i, also Figure S1 in Supporting Information S1). This is because the SOCAT data is primarily collected by ships, and the tough environment in the wintertime polar oceans (storms,



**Figure 2.** (a) The number of monthly,  $1^\circ \times 1^\circ$  grid cells with  $f\text{CO}_2$  for each year in Surface Ocean  $\text{CO}_2$  Atlas (SOCAT) v2022 and the standard deviation (std) of the seven  $\text{CO}_2$  flux data products in the Global Carbon Budget 2022 (Friedlingstein et al., 2022) for each year. (b) The standard deviation as a function of the number of monthly grid cells in SOCAT for each year.

darkness, and sea ice) makes access difficult. This seasonal skew may lead to an overemphasis on summertime  $\text{CO}_2$  measurements and less representation of wintertime processes (see e.g., Rustogi et al., 2023).

We randomly reduce the number of summer data sets in the Northern Ocean (June–August) and the Southern Ocean (December–February) to the same number as that in winter for each year (Figures 1g and 1i). Data sets in the Tropical Ocean are not adjusted (Figure 1h). The season in the Tropical Ocean is defined as the same as in the Southern Ocean. The winter-summer skew of the SOCAT data is minimized after this reduction in the summer data sets. We then re-grid the seasonally adjusted (i.e., summer data set reduced) SOCAT synthesis data set to generate another gridded data set for interpolation and flux estimation.

#### Experiment 4: Sensitivity of the flux to the addition of lower accuracy data

The  $f\text{CO}_2$  observations with a flag of E ( $f\text{CO}_{2\_E}$ , Figure S2 in Supporting Information S1) account for 17% of the total  $f\text{CO}_2$  data in SOCATv2023. All the  $f\text{CO}_2$  data in SOCAT are direct measurements, including the lower accuracy data (i.e., flag of E). SOCAT does not include any estimated  $f\text{CO}_2$  values from for example, pH and alkalinity, with a typical accuracy higher than  $10 \mu\text{atm}$  (e.g., Bushinsky et al., 2019; Williams et al., 2017). Therefore, the accuracy of the flag E data in SOCAT is better than that of the estimated  $f\text{CO}_2$ .

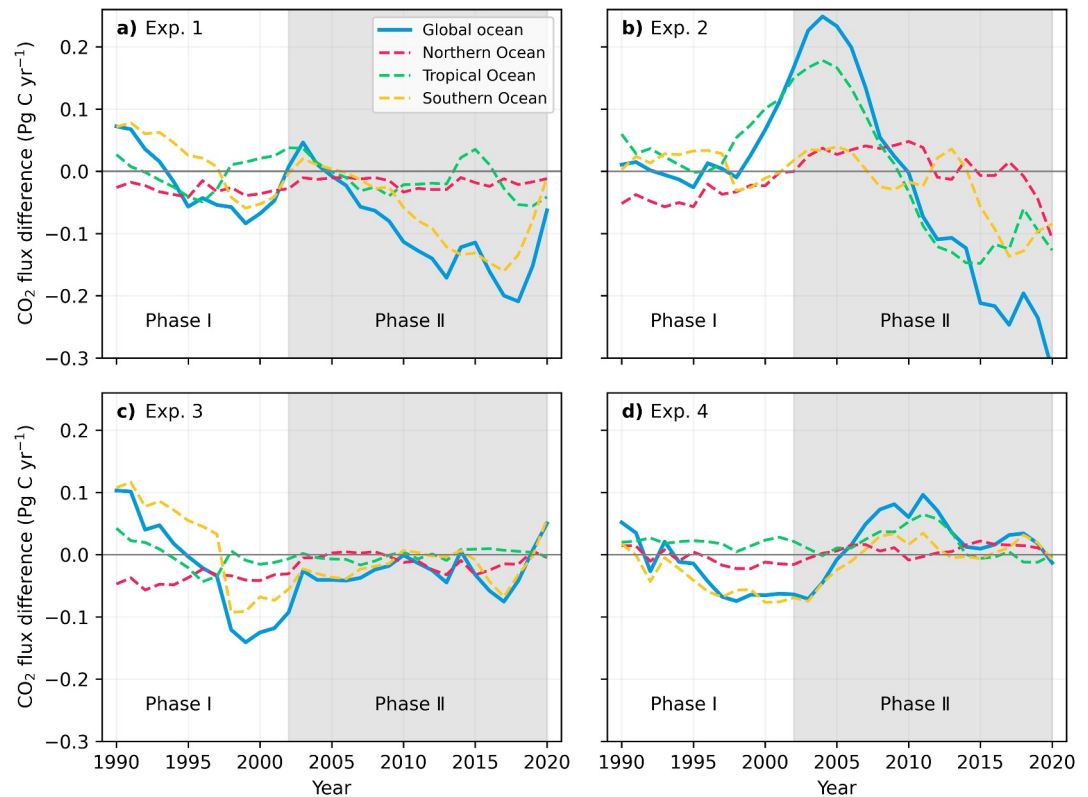
Here we include these lower-accuracy data into the flux estimate. The  $f\text{CO}_{2\_E}$  is not contained in the SOCAT synthesis data file but stored in an independent file. We first grid the  $f\text{CO}_{2\_E}$  data to a  $1^\circ$  by  $1^\circ$ , monthly resolution. We then replace the cells without  $f\text{CO}_2$  (in the SOCAT gridded product) by the co-located cells with  $f\text{CO}_{2\_E}$ . The new gridded  $f\text{CO}_2$  data set (SOCAT + Flag E, Figures 1j–1l) is further used for interpolation and flux calculation.

### 3. Results

#### 3.1. Spread of SOCAT-Based $\text{CO}_2$ Flux Estimates

From 2017 to 2021, the number of monthly grid cells with  $f\text{CO}_2$  has declined at a rate of about  $1,500 \text{ cells yr}^{-1}$ , similar to the sharpest rate of increase from 1999 to 2006. The Northern, Tropical, and Southern Oceans contribute 30%, 41%, and 29%, respectively, to the global decrease in the monthly grid cells with  $f\text{CO}_2$  (Figures 1j–1l). The Southern Ocean has the strongest relative decrease and its number of monthly grid cells with  $f\text{CO}_2$  in 2021 is only about 40% of that in 2017 (Figure 1l). The global coverage of grid cells with  $f\text{CO}_2$  in 2021 is the same as in 2005 (2.2%).

The number of  $1^\circ$  by  $1^\circ$ , monthly grid cells with  $f\text{CO}_2$  increased from  $\sim 1,000$  in 1990 to  $\sim 18,000$  in 2017, while the standard deviation of the SOCAT-based flux products decreased from  $0.48 \text{ Pg C yr}^{-1}$  to about  $0.15 \text{ Pg C yr}^{-1}$  over the same period (Figure 2a, Friedlingstein et al., 2022). In 2017, 3.5% of the monthly grid cells have  $f\text{CO}_2$  (or



**Figure 3.** Differences in global air-sea  $\text{CO}_2$  flux estimates for 1990–2020 between fluxes based on the four experimental data sets and those based on SOCATv2023. A neural network-based method has been used to interpolate Surface Ocean  $\text{CO}_2$  ATLAS (SOCAT)  $f\text{CO}_2$  to the global ocean. The blue, red, green, and yellow lines represent the flux in the global, Northern, Tropical, and Southern Oceans, respectively. Experimental fluxes based on SOCAT with (a) the number of data sets reduced to a similar number as in 2020 to simulate the recent decline in the availability; (b) the number of data sets reduced to a similar number as in 2000 to test long-term trends; (c) some summertime data sets removed to minimize the seasonal skew in the data; (d) additional lower accuracy data sets (flag of E).

$\sim 18,000$  grid cells), which corresponds to a standard deviation of the flux estimates of  $0.15 \text{ Pg C yr}^{-1}$  (Figure 2a). However, from 2017 to 2021, the number of monthly grid cells with  $f\text{CO}_2$  decreased by  $\sim 35\%$ , accompanied by an increase in the standard deviation to  $0.25 \text{ Pg C yr}^{-1}$  (65% relative increase). The negative correlation between the number of monthly grid cells with  $f\text{CO}_2$  and the standard deviation in the flux estimates demonstrates a strong dependence of the spread (or uncertainty) in the SOCAT-based  $\text{CO}_2$  flux estimates on the data availability in SOCAT (Figure 2b).

### 3.2. Sensitivity of the Annual Mean Flux to Data Availability and Distribution

To test the sensitivity of the annual mean air-sea  $\text{CO}_2$  flux estimate to the data availability and seasonal skew in the data, we carry out four experiments (see Section 2). The comparison between the SOCAT-based flux and the revised SOCAT-based flux is in Figure S3 of the Supporting Information S1. Here we show the flux sensitivity directly, as the flux difference ( $\Delta F$ ), defined as the experimental flux minus the original SOCAT-based flux. We use these metrics to characterize the flux sensitivity: (a) the mean and standard deviation of  $\Delta F$  from 1990 to 2002 (phase I) and from 2002 to 2020 (phase II); and (b) the trend of  $\Delta F$  in phase I and II. Here we choose 2002 as the separation year because the discrepancy in the flux trend between SOCAT-based products and models begins in 2002 (Friedlingstein et al., 2022).

In the first experiment, we test the impact of the recent decline in the SOCAT data on the annual mean  $\text{CO}_2$  flux estimates. Figure 3a shows that the  $\Delta F$  is small in the Northern and Tropical Oceans in both size and trends, but substantial in the Southern Ocean. Globally, the mean flux differences in both phases are on average less than  $0.1 \text{ Pg C yr}^{-1}$ , while the reduction of the data availability results in a 16% decrease in the positive flux trend in

phase I and an 18% increase in the negative trend in phase II (Table S1 in Supporting Information S1). These changes in the flux trend primarily come from the Southern Ocean (Figure 3a; Table S1 in Supporting Information S1). The sizeable  $\Delta F$  in phase I is noteworthy, especially as no data sets were removed during this period (Figures 1a–1c). The neural network trains for each biome on all the data collected from 1982 to 2020, and thus the reduction in data availability in later years also affects the training in the preceding years (Landschützer et al., 2013). In addition, repeat runs of the neural network training generate a slightly different annual mean flux (Figure S4 in Supporting Information S1, on average  $\sim \pm 0.05 \text{ Pg C yr}^{-1}$ ) as a result of the random selection of the training data and the validation data. Thus, the  $|\Delta F|$  values of less than  $0.05 \text{ Pg C yr}^{-1}$  can be ignored. In contrast, the flux trend does not have a notable difference between repeat runs (Figure S4 in Supporting Information S1).

Experiment two is designed to test if the SOCAT-based flux trend is sensitive to the limited data availability in the 1990s. Figure 3b shows that  $\Delta F$  is substantial in all three regions in both phases, especially in the Tropical Ocean. Global ocean  $\text{CO}_2$  uptake based on this experimental data set decreases more strongly (17%) in phase I and increases more quickly (55%) in phase II than that based on the original SOCAT data set (Table S1 in Supporting Information S1). Interestingly, although 72% of the SOCAT data sets were removed, the mean of  $\Delta F$  in both phases is small (less than  $0.1 \text{ Pg C yr}^{-1}$ , Table S1 in Supporting Information S1) but notable on an annual mean basis ( $-0.3$ – $0.25 \text{ Pg C yr}^{-1}$ ). We run two additional similar experiments (Exp. 2\_2 and Exp. 2\_3) to test if the result is sensitive to the randomness of the data reduction (see Figure S5 in Supporting Information S1). All three experiments show strong sensitivity of the global ocean  $\text{CO}_2$  flux trend to the data reduction, although the main source of the trend sensitivity comes from different regions (twice in the Tropical Ocean and once from the Southern Ocean).

The  $f\text{CO}_2$  observations in SOCAT have a seasonal skew with more data in summer than in winter (Figures 1g–1i). Figure 3c suggests that minimizing the winter–summer data skew reduces the increase in the trend of the  $\text{CO}_2$  flux in phase I by 35% but does not change the flux much in phase II (Table S1 in Supporting Information S1). In addition, this seasonal skew adjustment affects the seasonal variability of the Southern Ocean flux (Figure S6 in Supporting Information S1), with weaker  $\text{CO}_2$  uptake in winter than for the original SOCAT data set.

The addition of 7.1 million lower accuracy  $f\text{CO}_2$  values with a flag of E (Figure S2 in Supporting Information S1), equivalent to 17% of all SOCATv2023  $f\text{CO}_2$  values only increases the number of monthly grid cells with  $f\text{CO}_2$  by 2.1% (7565 grid cells from 1982 to 2020). This is because these lower accuracy data were mainly collected in coastal waters, while the  $1^\circ$  by  $1^\circ$ , monthly grid cells largely reflect open ocean areas. The addition of the flag E data slightly reduces the decreasing trend in the  $\text{CO}_2$  uptake in phase I and the increasing trend in phase II (Figure 3d; Table S1 in Supporting Information S1).

#### 4. Discussion and Conclusions

The oceans are regulating the global climate, and accurate quantification of the ocean  $\text{CO}_2$  uptake is crucial for climate mitigation and adaptation strategies. For this reason, scientists have established an annually updated system to quantify the ocean carbon sink alongside similar systems to assess  $\text{CO}_2$  emissions and the land carbon sink, which all contribute to the Global Carbon Budget (Friedlingstein et al., 2023). This system for the ocean includes surface ocean  $f\text{CO}_2$  observations, data synthesis, data analysis, and  $\text{CO}_2$  sink estimates (Guidi et al., 2020). The  $f\text{CO}_2$  observations collected by different research groups are gathered in SOCAT and quality controlled by experts (Bakker et al., 2016). The high-quality SOCAT data is then synthesized and gridded for analysis and global ocean  $\text{CO}_2$  flux estimates. SOCAT-based  $\text{CO}_2$  flux estimates inform climate negotiations part of the UN Framework Convention on Climate Change and will be crucial for the Global Greenhouse Gas Watch flagship initiative of the World Meteorological Organization (Bakker, Richard, et al., 2023).

Nonetheless, despite its importance, open ocean  $f\text{CO}_2$  data collection in SOCAT is in sharp decline. The number of  $f\text{CO}_2$  data sets per year decreased by 35% from 2017 to 2021, and the number of monthly grid cells with  $f\text{CO}_2$  decreased at a similar rate (33% from 2017 to 2021). The main reason for this decline is the lack of funding for data collection, certainly in some European countries. Running a  $f\text{CO}_2$  instrument on a ship (or other platform), processing the data, and submitting them to SOCAT requires funding. For instance, the UK-Caribbean Ship of Opportunity (SOOP) line (University of Exeter) ceased operation in 2019 for lack of funding, ending 17 years of data collection since 2002 (Schuster & Watson, 2007; Watson et al., 2009). Any observations that are not collected now, will not be available in the future (Wunsch et al., 2013). In addition, some data are being collected, but take several years to reach SOCAT because of the lack of data processing capacity.

Here we show that this recent decline in  $f\text{CO}_2$  observations has a remarkable impact on air-sea  $\text{CO}_2$  flux estimates (Figure 3a) and has significantly (65%) increased the spread in ocean  $\text{CO}_2$  uptake estimates for the years 2018–2021 in the Global Carbon Budget 2022 (Figure 2a). The strong negative correlation between the spread of the  $\text{CO}_2$  flux estimates and the SOCAT data availability (Figure 2b) demonstrates the importance of sustaining efforts in  $f\text{CO}_2$  data collection and highlights the risk to the accuracy of ocean  $\text{CO}_2$  sink estimates due to the recent alarming decline in the SOCAT data.

The data availability also affects the estimate of the flux trend (Figures 3a and 3b). The seven SOCAT-based products show a large spread in the flux trend before 2002 (Friedlingstein et al., 2023). SOCAT only contains very few data in the 1980s and the 1990s (less than 1% of the monthly grid cells have  $f\text{CO}_2$  values). Hauck, Gregor, et al. (2023) argued that this data sparse may lead to bias in the estimate of the SOCAT-based flux trend. Reducing the number of  $f\text{CO}_2$  data sets per year from 2001 to 2020 to that in 2000 increases the negative flux trend in this period by a factor of 1.5. This indicates that the SOCAT-based flux trend is very sensitive to data availability.

Seasonal skew in observational data is an inherent problem of the data collection effort. While it does not significantly affect the neural network-based  $\text{CO}_2$  flux on an annual basis from 2002 to 2020, it impacts the seasonal variability in the Southern Ocean flux estimates. This confirms the statement that our current understanding of the seasonality of the Southern Ocean carbon sink is hampered by the lack of high-accuracy  $f\text{CO}_2$  observations in winter (Hauck, Nissen, et al., 2023; Rustogi et al., 2023). Therefore, increasing the  $f\text{CO}_2$  data collection efforts in the Southern Ocean, especially in winter, should be a priority. Existing extra lower accuracy  $f\text{CO}_2$  data with a flag of E in SOCAT do not compensate for the impact of the decline in the  $f\text{CO}_2$  observations on the flux estimate.

It is worth noting that our experiments only show whether the  $\text{CO}_2$  flux is sensitive to the data availability or data distribution in SOCAT. The quantification of the sensitivity (Table S1 in Supporting Information S1) further depends on the randomly removed data sets (see Figure S5 in Supporting Information S1). Thus, the loss of some data sets might impact the air-sea  $\text{CO}_2$  flux estimates more significantly than that of others. Moreover, we only employ one neural network-based interpolation method for the experiment. Other interpolation method-based flux products (i.e., Rödenbeck et al., 2014) may react differently to data sparsity and thus our ability to determine the overall effect on the Global Carbon Budget is limited. We therefore suggest that further experiments should be extended to all available flux products (Friedlingstein et al., 2023) following the framework provided by this study.

In summary, SOCAT plays a crucial role in advancing our understanding of the contemporary ocean carbon uptake, but the recent decline in the  $f\text{CO}_2$  data availability threatens our ability to accurately estimate the ocean  $\text{CO}_2$  sink in both size and variability. Action is required to sustain and expand this observational network and its synthesis. The emerging uncrewed surface vehicles and platforms, such as surface moorings (Trowbridge et al., 2019), Saildrones (Sutton et al., 2021), and profiling floats (Williams et al., 2017), provide unprecedented opportunities for extensive  $f\text{CO}_2$  observations and estimates in the remote Southern Ocean in all seasons. Nevertheless, the accuracy of the  $f\text{CO}_2$  values based on some of these platforms may not be as high as that collected by ships, which highlights the need to continue the shipboard observations, including for provision of high-accuracy validation measurements.

## Data Availability Statement

Synthesis and gridded SOCATv2023: <https://www.socat.info/index.php/data-access/> (<https://doi.org/10.25921/r7xa-bt92>). Global Carbon Budget 2022: <https://globalcarbonbudgetdata.org/data-archive.html>.

## References

- Bakker, D. C. E., Alin, S. R., Bates, N., Becker, M., Feely, R. A., Gkritzalis, T., & Wimart-Rousseau, C. (2023). *Surface ocean  $\text{CO}_2$  Atlas database version 2023 (SOCATv2023)* (NCEI accession 0278913). NOAA National Centers for Environmental Information Dataset. <https://doi.org/10.25921/r7xa-bt92>
- Bakker, D. C. E., Pfeil, B., Landa, C. S., Metzl, N., O'Brien, K. M., Olsen, A., et al. (2016). A multi-decade record of high-quality  $f\text{CO}_2$  data in version 3 of the surface ocean  $\text{CO}_2$  Atlas (SOCAT). *Earth System Science Data*, 8(2), 383–413. <https://doi.org/10.5194/essd-8-383-2016>
- Bakker, D. C. E., Richard, S., Andrew, C., Michael, D., Thanos, G., Severino, I., et al. (2023). Case for SOCAT as an integral part of the value chain advising UNFCCC on ocean  $\text{CO}_2$  uptake. <https://socat.info/wp-content/uploads/2024/01/A-Case-for-SOCAT.pdf>

## Acknowledgments

The Surface Ocean  $\text{CO}_2$  Atlas (SOCAT) is an international effort, endorsed by the International Ocean Carbon Coordination Project (IOCCP), the Surface Ocean Lower Atmosphere Study (SOLAS) and the Integrated Marine Biogeochemistry and Ecosystem Research program (IMBER), to deliver a uniformly quality-controlled surface ocean  $\text{CO}_2$  database. D. C. E. Bakker leads SOCAT. The many researchers and funding agencies responsible for the collection of data and quality control are thanked for their contributions to SOCAT. For this work, Y. Dong has been supported by his family for his living expenses to complete this work. D. C. E. Bakker acknowledges the European Union's Horizon Europe research and innovation program under Grant agreement No 101056921 (project GreenFeedBack) and funding from UK Research and Innovation (UKRI) under Grant agreement 10040851 and Natural Environment Research Council (NERC)-funded PICCOLO Grants (R204504, NE/P021395/1). Views and opinions expressed are those of the authors only and do not necessarily reflect those of the European Union or UKRI. Neither the European Union nor UKRI can be held responsible for them.

- Bushinsky, S. M., Landschützer, P., Rödenbeck, C., Gray, A. R., Baker, D., Mazloff, M. R., et al. (2019). Reassessing Southern Ocean air-sea CO<sub>2</sub> flux estimates with the addition of biogeochemical float observations. *Global Biogeochemical Cycles*, *33*(11), 1370–1388. <https://doi.org/10.1029/2019GB006176>
- Denvil-Sommer, A., Gehlen, M., & Vrac, M. (2021). Observation system simulation experiments in the Atlantic Ocean for enhanced surface ocean pCO<sub>2</sub> reconstructions. *Ocean Science*, *17*(4), 1011–1030. <https://doi.org/10.5194/os-17-1011-2021>
- DeVries, T., Le Quééré, C., Andrews, O., Berthet, S., Hauck, J., Ilyina, T., et al. (2019). Decadal trends in the ocean carbon sink. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(24), 11646–11651. <https://doi.org/10.1073/pnas.1900371116>
- DeVries, T., Yamamoto, K., Wanninkhof, R., Gruber, N., Hauck, J., Müller, J. D., et al. (2023). Magnitude, trends, and variability of the global ocean carbon sink from 1985–2018. *Global Biogeochemical Cycles*, *37*(10), 1–32. <https://doi.org/10.1029/2023gb007780>
- Djeutchouang, L. M., Chang, N., Gregor, L., Vichi, M., & Monteiro, P. M. S. (2022). The sensitivity of pCO<sub>2</sub> reconstructions to sampling scales across a Southern Ocean sub-domain: A semi-idealized ocean sampling simulation approach. *Biogeosciences*, *19*(17), 4171–4195. <https://doi.org/10.5194/bg-19-4171-2022>
- Drugokencky, E., & Tans, P. (2023). *Trends in atmospheric carbon dioxide*. National Oceanic and Atmospheric Administration, Global Monitoring Laboratory (NOAA/GML). Retrieved from <https://gml.noaa.gov/ccgg/trends/global.html>
- Dong, Y., Bakker, D. C. E., Bell, T. G., Huang, B., Landschützer, P., Liss, P. S., & Yang, M. (2022). Update on the temperature corrections of global air-sea CO<sub>2</sub> flux estimates. *Global Biogeochemical Cycles*, *36*(9). <https://doi.org/10.1029/2022gb007360>
- Friedlingstein, P., O'Sullivan, M., Jones, M. W., Andrew, R. M., Gregor, L., Hauck, J., et al. (2022). Global carbon budget 2022. *Earth System Science Data*, *14*(11), 4811–4900. <https://doi.org/10.5194/essd-14-4811-2022>
- Friedlingstein, P., Sullivan, M. O., Jones, M. W., Andrew, R. M., Bakker, D. C. E., Hauck, J., et al. (2023). Global carbon budget 2023. *Earth System Science Data*, *15*(12), 5301–5369. <https://doi.org/10.18160/GCP-2023>
- 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), 1–14. <https://doi.org/10.1029/2020GB006788>
- Gruber, N., Bakker, D. C. E., DeVries, T., Gregor, L., Hauck, J., Landschützer, P., et al. (2023). Trends and variability in the ocean carbon sink. *Nature Reviews Earth & Environment*, *4*(2), 119–134. <https://doi.org/10.1038/s43017-022-00381-x>
- Guidi, L., Fernandez-Guerra, A., Canchaya, C., Curry, E., Fogliani, F., Irissou, J.-O., et al. (2020). Future science brief 6 of the European marine board. *Big Data in Marine Science*. <https://doi.org/10.5281/zenodo.3755792>
- Hauck, J., Gregor, L., Nissen, C., Patara, L., Hague, M., Mongwe, P., et al. (2023). The Southern Ocean carbon cycle 1985–2018: Mean, seasonal cycle, trends, and storage. *Global Biogeochemical Cycles*, *37*(11), e2023GB007848. <https://doi.org/10.1029/2023GB007848>
- Hauck, J., Nissen, C., Landschützer, P., Rödenbeck, C., Bushinsky, S., & Olsen, A. (2023). Sparse observations induce large biases in estimates of the global ocean CO<sub>2</sub> sink: An ocean model subsampling experiment. *Philosophical Transactions of the Royal Society A*, *381*(2249), 20220063. <https://doi.org/10.1098/rsta.2022.0063>
- IPCC. (2021). In V. Masson-Delmotte, P. Zhai, A. Pirani, S. L. Connors, C. Péan, 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. <https://doi.org/10.1017/9781009157896>
- Landschützer, P., Gruber, N., Bakker, D. C. E., Schuster, U., Nakaoka, S., Payne, M. R., et al. (2013). A neural network-based estimate of the seasonal to inter-annual variability of the Atlantic Ocean carbon sink. *Biogeosciences*, *10*(11), 7793–7815. <https://doi.org/10.5194/bg-10-7793-2013>
- Landschützer, P., Gruber, N., 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.aab262>
- Lee, H., Calvin, K., Dasgupta, D., Krinner, G., Mukherji, A., Thorne, P., et al. (2023). Climate change 2023: Synthesis report. Contribution of working groups I, II and III to the sixth assessment report of the intergovernmental panel on climate change. *IPCC*.
- Le Quééré, C., Peters, G. P., Andres, R. J., Andrew, R. M., Boden, T. A., Ciais, P., et al. (2014). Global carbon budget 2013. *Earth System Science Data*, *6*(1), 235–263. <https://doi.org/10.5194/essd-6-235-2014>
- Ostle, C., Landschützer, P., Edwards, M., Johnson, M., Schmidtko, S., Schuster, U., et al. (2022). Multidecadal changes in biology influence the variability of the North Atlantic carbon sink. *Environmental Research Letters*, *17*(11), 114056. <https://doi.org/10.1088/1748-9326/ac9ecf>
- Pfeil, B., Olsen, A., Bakker, D. C. E., Hankin, S., Koyuk, H., Kozyr, A., et al. (2013). A uniform, quality controlled Surface Ocean CO<sub>2</sub> Atlas (SOCAT). *Earth System Science Data*, *5*(1), 125–143. <https://doi.org/10.5194/essd-5-125-2013>
- Ritter, R., Landschützer, P., Gruber, N., Fay, A. R., Iida, Y., Jones, S., et al. (2017). Observation-based trends of the Southern Ocean carbon sink. *Geophysical Research Letters*, *44*(24), 12339–12348. <https://doi.org/10.1002/2017GL074837>
- Rödenbeck, C., Bakker, D. C. E., Metzl, N., Olsen, A., Sabine, C., Cassar, N., et al. (2014). Interannual sea-air CO<sub>2</sub> flux variability from an observation-driven ocean mixed-layer scheme. *Biogeosciences*, *11*(17), 4599–4613. <https://doi.org/10.5194/bg-11-4599-2014>
- Rödenbeck, C., Keeling, R. F., Bakker, D. C. E., Metzl, N., Olsen, A., Sabine, C., & Heimann, M. (2013). Global surface-ocean pCO<sub>2</sub> and sea-air CO<sub>2</sub> flux variability from an observation-driven ocean mixed-layer scheme. *Ocean Science*, *9*(2), 193–216. <https://doi.org/10.5194/os-9-193-2013>
- Rodgers, K. B., Schwinger, J., Fassbender, A. J., Landschützer, P., Yamaguchi, R., Frenzel, H., et al. (2023). Seasonal variability of the surface ocean carbon cycle: A synthesis. *Global Biogeochemical Cycles*, *37*(9), 1–34. <https://doi.org/10.1029/2023GB007798>
- Rustogi, P., Landschützer, P., Brune, S., & Baehr, J. (2023). The impact of seasonality on the annual air-sea carbon flux and its interannual variability. *Npj Climate and Atmospheric Science*, *6*(1), 66. <https://doi.org/10.1038/s41612-023-00378-3>
- Sabine, C. L., Hankin, S., Koyuk, H., Bakker, D. C. E., Pfeil, B., Olsen, A., et al. (2013). Surface ocean CO<sub>2</sub> Atlas (SOCAT) gridded data products. *Earth System Science Data*, *5*(1), 145–153. <https://doi.org/10.5194/essd-5-145-2013>
- Schuster, U., & Watson, A. J. (2007). A variable and decreasing sink for atmospheric CO<sub>2</sub> in the North Atlantic. *Journal of Geophysical Research*, *112*(C11), C11006. <https://doi.org/10.1029/2006JC003941>
- Sutton, A. J., Williams, N. L., & Tilbrook, B. (2021). Constraining Southern Ocean CO<sub>2</sub> flux uncertainty using uncrewed surface vehicle observations. *Geophysical Research Letters*, *48*(3), 1–9. <https://doi.org/10.1029/2020GL091748>
- Trowbridge, J., Weller, R., Kelley, D., Dever, E., Plueddemann, A., Barth, J. A., & Kawka, O. (2019). The ocean observatories initiative. *Frontiers in Marine Science*, *6*, 434166. <https://doi.org/10.3389/fmars.2019.00074>
- Wanninkhof, R. (2014). Relationship between wind speed and gas exchange over the ocean revisited. *Limnology and Oceanography: Methods*, *12*(6), 351–362. <https://doi.org/10.4319/lom.2014.12.351>
- Watson, A. J., Schuster, U., Bakker, D. C. E., Bates, N. R., Corbière, A., González-Dávila, M., et al. (2009). Tracking the variable North Atlantic sink for atmospheric CO<sub>2</sub>. *Science*, *326*(5958), 1391–1393. <https://doi.org/10.1126/science.1177394>
- 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<sub>2</sub> flux are consistent with ocean carbon inventory. *Nature Communications*, *11*(1), 1–6. <https://doi.org/10.1038/s41467-020-18203-3>



- Weiss, R. F. (1974). Carbon dioxide in water and seawater: The solubility of a non-ideal gas. *Marine Chemistry*, 2(3), 203–215. [https://doi.org/10.1016/0304-4203\(74\)90015-2](https://doi.org/10.1016/0304-4203(74)90015-2)
- Williams, N. L., Juranek, L. W., Feely, R. A., Johnson, K. S., Sarmiento, J. L., Talley, L. D., et al. (2017). Calculating surface ocean  $p\text{CO}_2$  from biogeochemical Argo floats equipped with pH: An uncertainty analysis. *Global Biogeochemical Cycles*, 31(3), 591–604. <https://doi.org/10.1002/2016GB005541>
- Woolf, D. K., Land, P. E., Shutler, J. D., Goddijn-Murphy, L. M., & Donlon, C. J. (2016). On the calculation of air-sea fluxes of  $\text{CO}_2$  in the presence of temperature and salinity gradients. *Journal of Geophysical Research: Oceans*, 121(2), 1229–1248. <https://doi.org/10.1002/2015JC011427>
- Wunsch, C., Schmitt, R. W., & Baker, D. J. (2013). Climate change as an intergenerational problem. *Proceedings of the National Academy of Sciences of the United States of America*, 110(12), 4435–4436. <https://doi.org/10.1073/pnas.1302536110>

## References From the Supporting Information

- Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., et al. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, 146(730), 1999–2049. <https://doi.org/10.1002/qj.3803>
- Ho, D. T., Law, C. S., Smith, M. J., 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. *Geophysical Research Letters*, 33(16), L16611. <https://doi.org/10.1029/2006GL026817>
- Landschützer, P., Gruber, N., Bakker, D. C. E., & Schuster, U. (2014). Recent variability of the global ocean carbon sink. *Global Biogeochemical Cycles*, 28(9), 927–949. <https://doi.org/10.1002/2014GB004853>
- Merchant, C. J., Embury, O., Bulgin, C. E., Block, T., Corlett, G. K., Fiedler, E., et al. (2019). Satellite-based time-series of sea-surface temperature since 1981 for climate applications. *Scientific Data*, 6(1), 1–18. <https://doi.org/10.1038/s41597-019-0236-x>
- Naegler, T. (2009). Reconciliation of excess  $^{14}\text{C}$ -constrained global  $\text{CO}_2$  piston velocity estimates. *Tellus B: Chemical and Physical Meteorology*, 61(2), 372–384. <https://doi.org/10.1111/j.1600-0889.2008.00408.x>