

## Improved winter data coverage of the Southern Ocean CO<sub>2</sub> sink from extrapolation of summertime observations

Neill Mackay <sup>1</sup>, Andrew J. Watson <sup>1</sup>, Parvada Suntharalingam<sup>2</sup>, Zhaohui Chen<sup>2</sup> & Peter Landschützer <sup>3,4</sup>

The Southern Ocean is an important sink of anthropogenic CO<sub>2</sub>, but it is among the least well-observed ocean basins, and consequentially substantial uncertainties in the CO<sub>2</sub> flux reconstruction exist. A recent attempt to address historically sparse wintertime sampling produced ‘pseudo’ wintertime observations of surface pCO<sub>2</sub> using subsurface summertime observations south of the Antarctic Polar Front. Here, we present an estimate of the Southern Ocean CO<sub>2</sub> sink that combines a machine learning-based mapping method with an updated set of pseudo observations that increases regional wintertime data coverage by 68% compared with the historical dataset. Our results confirm the suggestion that improved winter coverage has a modest impact on the reconstruction, slightly strengthening the uptake trend in the 2000s. After also adjusting for surface boundary layer temperature effects, we find a 2004–2018 mean sink of  $-0.16 \pm 0.07$  PgC yr<sup>-1</sup> south of the Polar Front and  $-1.27 \pm 0.23$  PgC yr<sup>-1</sup> south of 35°S, consistent with independent estimates from atmospheric data.

<sup>1</sup>Faculty of Environment, Science and Economy, University of Exeter, Exeter, UK. <sup>2</sup>School of Environmental Sciences, University of East Anglia, Norwich, UK. <sup>3</sup>Max Planck Institute for Meteorology, 20146 Hamburg, Germany. <sup>4</sup>Flanders Marine Institute (VLIZ), InnovOceanSite, 8400 Ostend, Belgium. ✉email: [n.mackay@exeter.ac.uk](mailto:n.mackay@exeter.ac.uk)

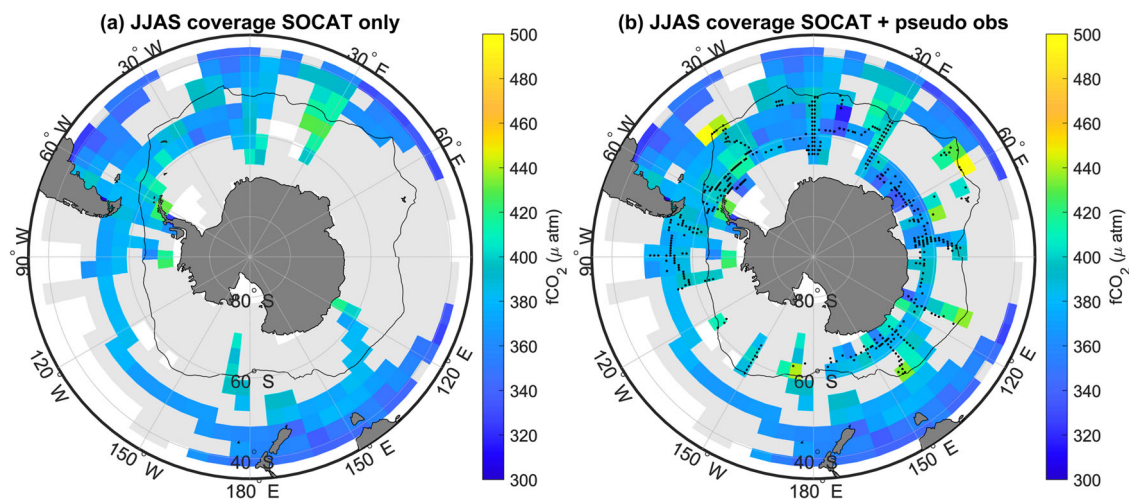
The ocean is an important sink of anthropogenic CO<sub>2</sub>, having absorbed 23% of all man-made emissions in 2019<sup>1</sup>, with the Southern Ocean south of 35°S accounting for around 40% of this global ocean sink<sup>2</sup>. This outsized role in the climate system means that understanding the Southern Ocean carbon sink and its variability are critical for climate assessments and for the Global Carbon Budget<sup>1</sup> that contributes to the Intergovernmental Panel on Climate Change report<sup>3</sup>. Estimates of air-sea CO<sub>2</sub> fluxes from models, observation-based data products and observations show the largest disagreement in the Southern Ocean, in particular with regard to low frequency variations<sup>4</sup>. Given the lack of observational coverage in this remote and harsh ocean region, this disagreement does not come as a surprise. The Surface Ocean CO<sub>2</sub> Atlas (SOCAT)<sup>5</sup> is the largest compilation of all surface CO<sub>2</sub> observations, but large areas of the Southern Ocean remain sparsely or unsampled in winter, particularly at higher latitudes (Fig. 1a and S1). To fill this data void, ocean carbon sink estimates are produced by interpolating in-situ measurements using a variety of techniques including state-of-the-art machine learning methods<sup>6–9</sup>, and air-sea CO<sub>2</sub> fluxes are then computed by combining the mapped values with atmospheric CO<sub>2</sub> data using a simple bulk gas transfer parameterization<sup>10</sup>. Despite the recent advancements in gap-filling methods, however, data sparsity continues to be among the largest uncertainties in the flux reconstructions, particularly in the Southern Ocean<sup>4</sup>.

Recognizing the need for year-round observations to monitor this essential carbon sink, since 2014 the observational coverage of the Southern Ocean has rapidly improved thanks to the advent of autonomous platforms including floats<sup>11</sup> and uncrewed surface vehicles<sup>12</sup>. Notwithstanding the substantial improvement in coverage they bring, these new observations cannot address historical sparsity and are therefore limited in their ability to constrain estimates of past multiyear variability. In an attempt to resolve this, a recent study<sup>13</sup> used summertime subsurface observations of dissolved inorganic carbon (DIC) south of the Antarctic Polar Front (APF) from the Global Ocean Data Analysis Project (GLODAP)<sup>14,15</sup> to extrapolate ‘pseudo observations’ of wintertime surface pCO<sub>2</sub>, boosting the coverage from 2004 onwards. They found that the pseudo observations increased winter outgassing, but their method did

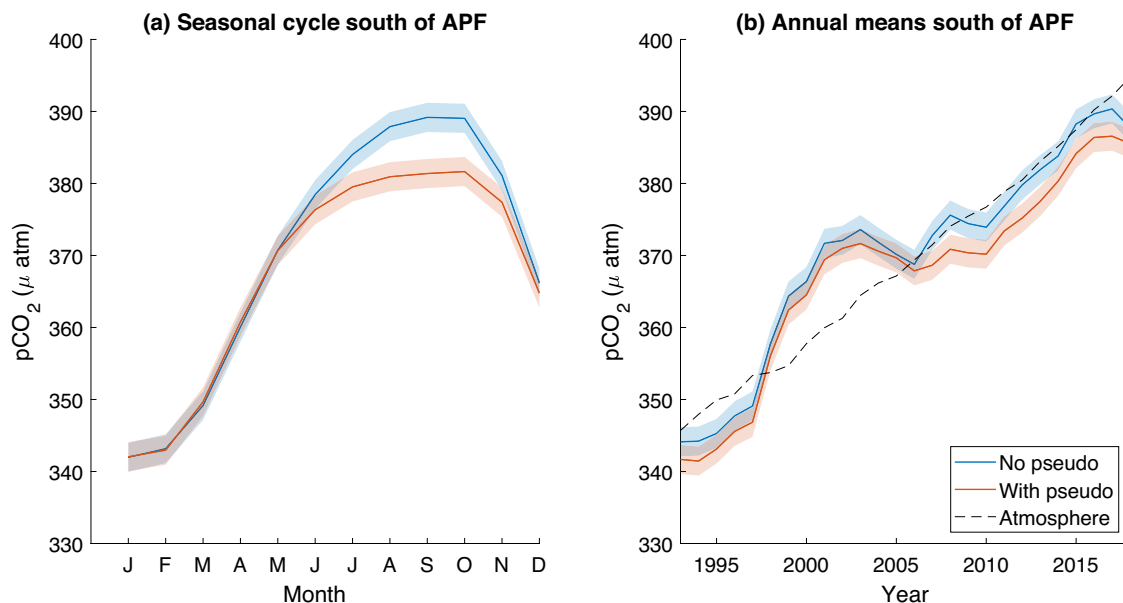
not reproduce a strong reversal of the trend from an increasing to a decreasing sink around 2011 suggested by other studies<sup>7,16</sup>. In this work, we build on this novel constraint in two ways: first by switching from a simple multiple linear regression (MLR) for gap-filling the data to a more sophisticated neural network-based approach<sup>17</sup>; and second by increasing the number of pseudo observations from 760 to 798 using a more recent version of GLODAP<sup>18</sup>. We further validate the full set of pseudo observations against a data-assimilating biogeochemical ocean model<sup>19</sup>, and combine our mapped fCO<sub>2</sub> product with a gas transfer parameterisation<sup>10</sup> to produce air–sea CO<sub>2</sub> flux estimates for the Southern Ocean from 1993 to 2018 (see Methods section).

When binned monthly and onto a 4° latitude by 6° longitude grid (roughly equivalent to the 400 km decorrelation length scale for pCO<sub>2</sub><sup>20</sup>), our pseudo observations increase wintertime (June–September) coverage south of the APF by 68% for the period 2004–2018 compared with using only the direct observations compiled in the SOCAT database<sup>5</sup> (see Fig. 1). However, the pseudo observations are assumed to represent conditions in September (see Methods section), and as such a majority of them are in locations that would be under sea ice, which diminishes their direct influence on the air–sea flux. Instead, it is the modification by the pseudo observations of relationships between driver variables and surface fCO<sub>2</sub> established in the gap-filling step that refines our estimate of the carbon sink. An illustration of the differences between the estimate of surface CO<sub>2</sub> from this study and an estimate using the same set of pseudo observations combined with the MLR gap-filling method of previously employed<sup>13</sup> is shown in Supplementary Fig. S1.

Furthermore, recent research<sup>21</sup> highlighted the need to make a correction to the fCO<sub>2</sub> data to account for temperature gradients in the atmosphere–ocean mass boundary layer (MBL) before calculating the air–sea flux. This altered the 1992–2018 global mean ocean CO<sub>2</sub> sink by 0.8–0.9 PgC yr<sup>−1</sup> (ref. 22), yet no regional correction for the Southern Ocean exists. Here we apply a similar temperature correction, and compare our estimate of the Southern Ocean carbon sink with results from an atmospheric inversion that relies on atmospheric observations of CO<sub>2</sub> from surface sites in combination with model simulations of atmospheric CO<sub>2</sub> transport<sup>23</sup> (see Methods section).



**Fig. 1 Southern ocean fCO<sub>2</sub> data coverage.** **a** Colour shows wintertime (June–September) observations of sea surface fCO<sub>2</sub> (in micro-atmosphere, µatm) from SOCAT south of 35°S from 2004 to 2018, binned and averaged onto a 4° latitude by 6° longitude grid. **b** As **a** but with the addition of wintertime pseudo fCO<sub>2</sub> observations constructed for the same period. Light grey shading shows the coverage of all non-wintertime SOCAT fCO<sub>2</sub> observations for the same years on the same grid, and the black dots are the locations of the pseudo observations on the 1° × 1° grid used to train the SOM-FFN. The black lines show the mean position of the polar front from a published product<sup>62</sup>.



**Fig. 2** Mean pCO<sub>2</sub> from SOM-FFN reconstructions. **a** 1993–2018 mean seasonal cycle and **b** annual mean surface pCO<sub>2</sub> (in micro-atmosphere, µ atm) south of the Antarctic Polar Front from the SOM-FFN with and without pseudo observations in the training data (solid lines). Each line represents an ensemble mean, and the shaded areas are the 1-σ uncertainties (see Methods section). The black dashed line shows the annual mean atmospheric pCO<sub>2</sub> above the same ocean area, calculated from the NOAA ESRL product<sup>50,51</sup>.

## Results

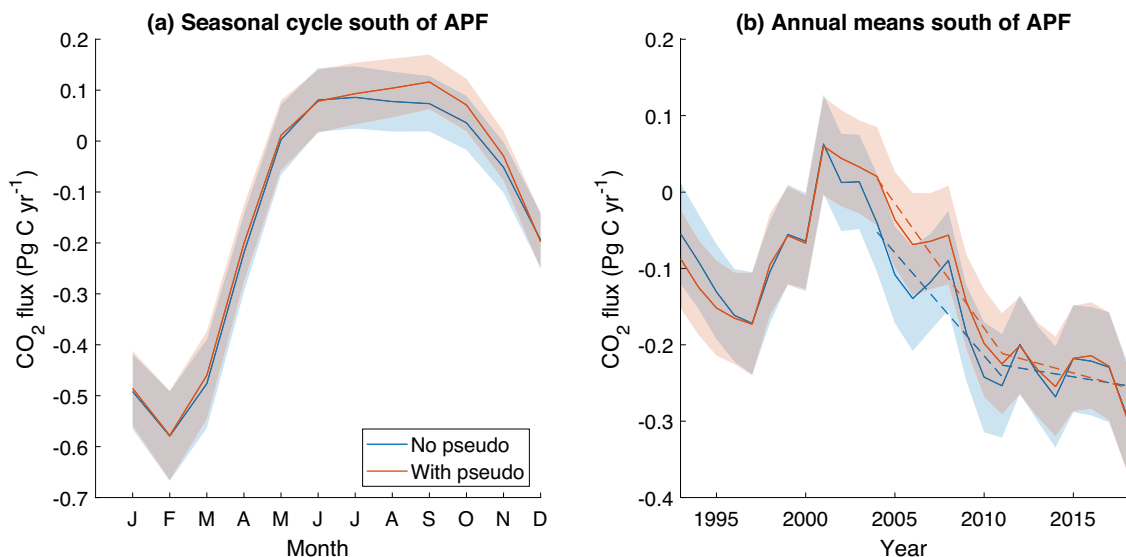
**Surface pCO<sub>2</sub> south of the Antarctic Polar Front.** Figure 2 shows the mean seasonal cycle and annual means south of the APF of surface pCO<sub>2</sub> mapped using a Self-Organising Map Feed Forward Neural Network technique (SOM-FFN; see Methods section). The winter peak of the seasonal cycle of pCO<sub>2</sub> in the region is reduced due to the pseudo observations. The phase of the seasonal cycle of pCO<sub>2</sub> is unaffected by their inclusion, but its magnitude reduces from 47 µatm to 40 µatm. The seasonal cycle of pCO<sub>2</sub> is anticorrelated with temperature, and instead most closely resembles the non-thermal drivers mixed layer depth and sea surface salinity, which have their peak values in September and October (see Supplementary Fig. S2). This suggests the seasonal cycle of pCO<sub>2</sub> is DIC-driven in this region, with deeper mixed layers in wintertime stirring up DIC-rich waters from below causing an outgassing tendency that overcomes the uptake tendency caused by surface cooling.

The long-term trend follows the atmospheric pCO<sub>2</sub> trend (Fig. 2b), but with some strong variability in the 2000s in particular that is not reduced by the inclusion of pseudo observations. From 2007 onwards, the pseudo observations cause only a slight divergence in the two solid lines, and a slightly weaker upward trend such that the surface pCO<sub>2</sub> estimate does not keep pace with the atmosphere.

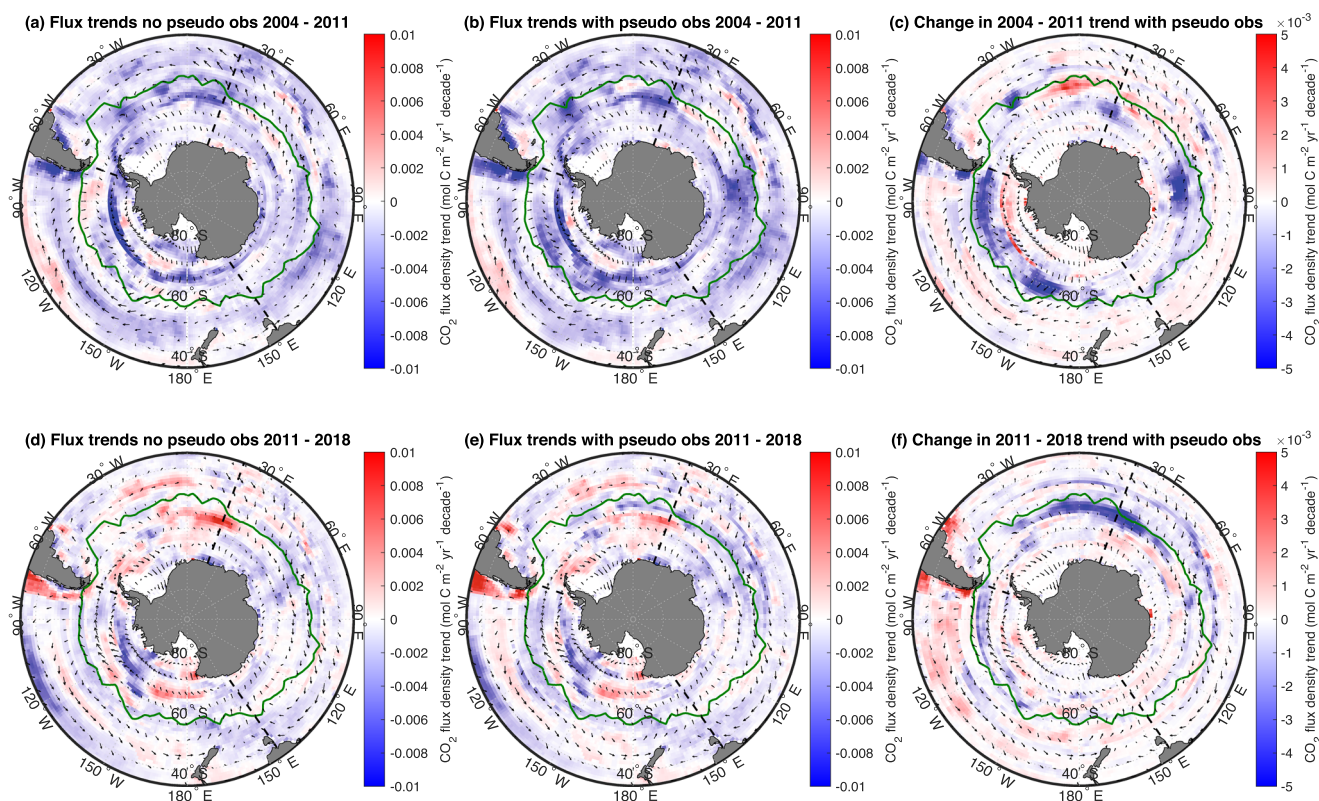
**Air-sea CO<sub>2</sub> fluxes south of the Antarctic Polar Front.** The reduction of the winter peak of the seasonal cycle due to the pseudo observations seen on Fig. 2a would alone tend to increase the estimated CO<sub>2</sub> uptake by the ocean. However, the reduction in pCO<sub>2</sub> occurs below the sea ice (see Supplementary Fig. S3), and, consequently, this signal has little contribution to the air-sea flux: away from sea ice, the pseudo observations tend to increase pCO<sub>2</sub> and this effect just dominates the air-sea flux south of the APF (Figs. S3 and 3a). The peak winter outgassing increases fractionally from 0.09 PgC yr<sup>-1</sup> to 0.12 PgC yr<sup>-1</sup>, and is shifted from July to September attributed to the data interpolation method; meanwhile the magnitude of the flux seasonal cycle increases from 0.66 to 0.69 PgC yr<sup>-1</sup>.

The fluxes on Fig. 3 with and without the pseudo observations are not distinguishable within the uncertainties, but the central estimate of the annual mean flux is shifted upwards slightly in the period 2002 to 2011. The pseudo observations begin in 2004 due to their method of calculation (see Methods section), so their influence diminishes before then; after 2011 the relative increase in coverage they provide also reduces due to improvements in direct fCO<sub>2</sub> measurement coverage (see Supplementary Fig. S4). There is only a slight impact on the multiyear variability of the fluxes over the period covered by the pseudo observations: both lines on Fig. 3b show an increase in the sink from 2004 to 2011 followed by a stagnation. Without the pseudo observations the sink trend south of the APF from 2004 to 2011 is  $-0.27 \text{ PgC yr}^{-1} \text{ decade}^{-1}$ , and from 2011 to 2018 is  $-0.04 \text{ PgC yr}^{-1} \text{ decade}^{-1}$ . When the pseudo observations are added, the trends over the same two periods are  $-0.33 \text{ PgC yr}^{-1} \text{ decade}^{-1}$  and  $-0.06 \text{ PgC yr}^{-1} \text{ decade}^{-1}$ , respectively (linear trends are shown as dashed lines on Fig. 3). For the period between 2004, when the pseudo observations begin, to 2011, where their influence largely vanishes, they reduce the mean sink by 34% from 0.15 PgC yr<sup>-1</sup> to 0.10 PgC yr<sup>-1</sup>.

**Spatial patterns of air-sea CO<sub>2</sub> flux trends.** The decadal trends shown on Fig. 3b support the view of a weakening sink in the 1990s<sup>24</sup>, followed by a reinvigoration in the 2000s<sup>25</sup> and a subsequent stagnation in at least the early part of the 2010s<sup>26</sup>. We now examine the spatial variation of the decadal trends from 2004 onwards, and the impact of the pseudo observations on those trends, on Fig. 4. While direction and spatial patterns are largely unaltered, we observe a regional refinement of the trends. In 2004–2011 (Fig. 4a, b), the trend is towards increasing uptake almost everywhere across the Southern Ocean, and the pseudo observations even strengthen the trend south of the zonal wind maximum in all sectors, with the largest effect in the Pacific (Fig. 4c). In 2011–2018, the trend is patchier with areas of increasing uptake and areas of increasing outgassing. The pseudo observations strengthen an outgassing trend in the Pacific lower



**Fig. 3** Air-sea  $\text{CO}_2$  fluxes south of the Antarctic Polar Front. **a** 1993–2018 mean seasonal cycle and **b** annual mean air-sea  $\text{CO}_2$  fluxes (negative into the ocean) south of the Antarctic Polar Front based on sea surface  $\text{fCO}_2$  from the SOM-FFN with and without pseudo observations in the training data. Values are Petagrams of carbon per year ( $\text{Pg C yr}^{-1}$ ). Each solid line represents an ensemble mean, and the shaded areas are the  $1\text{-}\sigma$  uncertainties (see Methods section). The dashed lines on **b** are linear fits for the 2004–2011 and 2011–2018 periods.



**Fig. 4** Mapped  $\text{CO}_2$  flux trends. Colours show air-sea  $\text{CO}_2$  flux trends (negative is increasing uptake) and arrows show wind trends, for the 2004–2011 period (top row) and 2011–2018 period (bottom row). Values are moles of carbon per metre squared per year per decade ( $\text{mol C m}^{-2} \text{yr}^{-1} \text{decade}^{-1}$ ). Panels **a** and **d** show  $\text{CO}_2$  flux trends when pseudo observations are not used, panels **b** and **e** show the trends when the pseudo observations are included, and panels **c** and **f** show the difference in  $\text{CO}_2$  flux trends due to the pseudo observations (note the different colour scale for the difference plots). The solid green line on all panels shows the latitude of the 2004–2018 mean zonal wind maximum based on CCMP winds<sup>46</sup>, and the black dashed lines mark ocean basin boundaries.

latitudes, and weaken a region of outgassing trend just south of the zonal wind maximum near  $20^\circ\text{E}$  (Fig. 4e, f).

The weakening of the Southern Ocean carbon sink in the 1990s has been explained by a southward shift of westerly winds associated with an increasing positive index of the Southern

Annual Mode, which caused increased upwelling of natural DIC and consequent outgassing of  $\text{CO}_2$ <sup>27</sup>. The reinvigoration from 2002 to 2011 has been attributed to a zonally asymmetric atmospheric circulation that drove contrasting patterns of wind and SST change in the different ocean basins<sup>25,26</sup>, and the



subsequent stagnation to regional shifts in sea level pressure linked to its zonal wavenumber 3 and related changes in winds<sup>26</sup>. Other explanations include changes in external forcing<sup>28</sup>, and changes in the strength of the upper ocean overturning circulation<sup>29</sup>. Here we are unable to further elaborate on the cause of the decadal variations, however, we can speculate as to whether the addition of our pseudo observations supports the previously suggested mechanisms. For example, the enhanced 2004 to 2011 uptake trend at high latitudes (Fig. 4c) coincides with a weakening trend in the westerly winds that would likely be associated with changes in the physical circulation. This appears consistent with suggestions that winter variability driven by changes in mixing and stratification controls the longer term variability of the Southern Ocean carbon sink<sup>30</sup>. This is also supported by our finding that the seasonal cycle of pCO<sub>2</sub> in the region closely resembles that of the mixed layer depth (Supplementary Fig. S2).

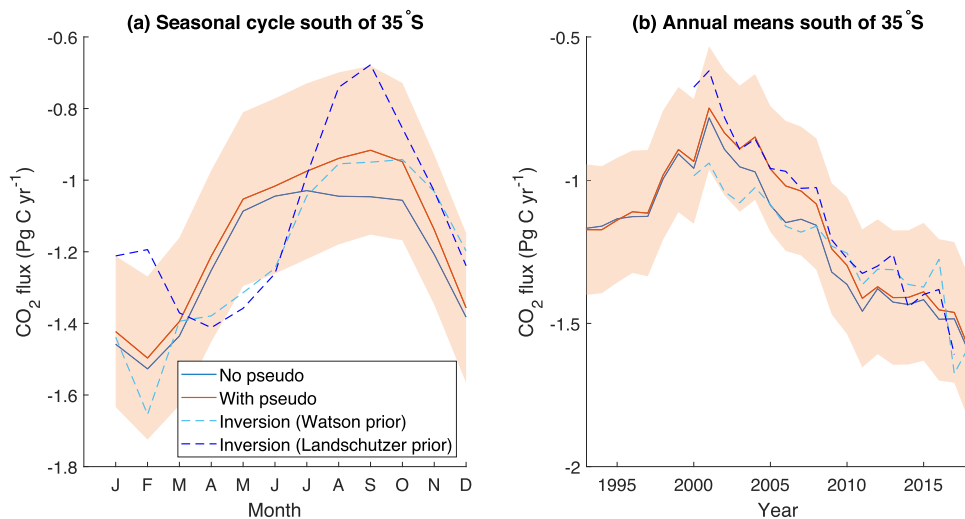
**Comparison with atmospheric inversions.** We have so far examined the mean fluxes for the region of the Southern Ocean southwards of the APF, since this is where we have added information in the form of our pseudo observations to aid in the estimation of the carbon sink. To place the results in context, we show the sink for the whole Southern Ocean south of 35°S, on Fig. 5. The results are similar to those for the high latitudes on Fig. 3, with the pseudo observations causing a small reduction in uptake over the winter period (Fig. 5a) and in the mid-late 2000s (Fig. 5b). The central estimate of the 2004 to 2011 mean sink reduces by 8% from  $-1.21 \text{ PgC yr}^{-1}$  without the pseudo observations to  $-1.11 \text{ PgC yr}^{-1}$  when they are included.

Corrections applied to account for MBL effects have increased the sink south of 35°S by  $\sim 0.3 \text{ PgC yr}^{-1}$  compared with estimates from fluxes where the corrections were not applied to the fCO<sub>2</sub> data driving the SOM-FFN mapping (see Supplementary Fig. S5); this offset is in line with a previous estimate for the global ocean<sup>21</sup>. The combination of the reduction due to the pseudo observations offset by the increase due to the MBL corrections results in an estimated 2004–2018 mean sink south of the APF of  $-0.16 \pm 0.07 \text{ PgC yr}^{-1}$ , and south of 35°S of  $-1.27 \pm 0.23 \text{ PgC yr}^{-1}$ . Our results are broadly consistent with two atmospheric inversions (Fig. 5 dashed lines),

which are based on a set of independent measurements, and with a recent study that estimated the Southern Ocean sink using aircraft data<sup>31</sup> (see Supplementary Fig. S6). We note here that the MBL correction is crucial to bringing our results in line with the atmospheric estimates, with the pseudo observations only improving the agreement with one of the atmospheric inversions on Fig. 5b, and making almost no difference to the comparison with aircraft data on Fig. S6.

## Discussion

We have presented an estimate of the Southern Ocean CO<sub>2</sub> sink that applied a sophisticated pCO<sub>2</sub> gap-filling technique to a dataset with boosted wintertime coverage in the form of pseudo observations extrapolated from subsurface summertime observations. Our work builds on the results of an earlier study<sup>13</sup> that attempted to estimate the carbon sink and its variability south of the APF with the benefit of pseudo observations constructed using the same methods, but with the necessary step of gap-filling ocean surface CO<sub>2</sub> data having been achieved using a multiple linear regression, rather than the more sophisticated SOM-FFN machine learning method used here. The earlier study estimated the long-term mean sink from 2004 to 2017 as  $-0.02 \pm 0.02 \text{ PgC yr}^{-1}$  south of the APF. After accounting for a downward shift of  $0.10 \text{ PgC yr}^{-1}$  for that region due to the MBL corrections applied in this study, our estimate of  $-0.16 \pm 0.07 \text{ PgC yr}^{-1}$  agrees well with the earlier result. Qualitatively the variability over the same period is also quite similar between the two studies: when split at 2011 and run separately over two time periods, the MLR showed an increasing sink from 2004 to 2011 followed by a flattening/slight decrease (their Fig. D1). It also showed, as we do here, that the introduction of pseudo observations increases the downward trend over the 2004–2011 period (Fig. 3b). Quantitatively, the increase in the sink of  $0.32 \text{ PgC yr}^{-1}$  from the MLR implies a trend of  $\sim -0.46 \text{ PgC yr}^{-1} \text{ decade}^{-1}$  which is larger than our trend of  $-0.33 \text{ PgC yr}^{-1} \text{ decade}^{-1}$  using the SOM-FFN. The difference could be explained by a number of factors, including the different gap-filling methods, the longer training dataset used in the current study (1993–2018 compared to 2004–2017), a larger set of driver variables (see Table 1; the MLR study used only temperature, salinity, mixed layer depth and atmospheric



**Fig. 5 Air-sea CO<sub>2</sub> fluxes for the Southern Ocean.** **a** 2000–2017 mean seasonal cycle and **b** annual mean air-sea CO<sub>2</sub> fluxes (negative into the ocean) for the Southern Ocean south of 35°S based on sea surface fCO<sub>2</sub> from the SOM-FFN with and without pseudo observations. Values are Petagrams of carbon per year (Pg C yr<sup>-1</sup>). Each line represents an ensemble mean, and the shaded area is the 1-σ uncertainty on the SOM-FFN result with pseudo observations (see Methods). Also plotted are the same quantities derived from an atmospheric inversion using two different priors (dashed lines; see Methods section). Note that unlike Fig. 3, the mean seasonal cycle is plotted for a shorter time period, for consistency with the atmospheric inversions.

**Table 1 Sources of data.**

Variable	Used for	Data source	Product and reference
Sea surface fCO <sub>2</sub>	SOM-FFN	<a href="https://www.socat.info/index.php/version-2020/">https://www.socat.info/index.php/version-2020/</a>	SOCAT V2020 <sup>5</sup>
Sea surface temperature	SOM-FFN, fCO <sub>2</sub> temperature correction and air-sea flux	<a href="https://www.emc.ncep.noaa.gov/research/cmb/sst_analysis/">https://www.emc.ncep.noaa.gov/research/cmb/sst_analysis/</a>	NOAA Optimum Interpolation product v2 <sup>36</sup>
Sea surface salinity	SOM-FFN	<a href="https://www.metoffice.gov.uk/hadobs/en4/download-en4-2-1.html">https://www.metoffice.gov.uk/hadobs/en4/download-en4-2-1.html</a>	Hadley Centre EN4 version EN4.2.1 <sup>56</sup>
Sea ice concentration	SOM-FFN	<a href="https://www.metoffice.gov.uk/hadobs/hadist/">https://www.metoffice.gov.uk/hadobs/hadist/</a>	HadISST <sup>57</sup>
Mixed layer depth	SOM-FFN	<a href="https://resources.marine.copernicus.eu/product-detail/MULTIOBS_GLO_PHY_TSUV_3D_MYNRT_015_012/INFORMATION">https://resources.marine.copernicus.eu/product-detail/MULTIOBS_GLO_PHY_TSUV_3D_MYNRT_015_012/INFORMATION</a>	CMEMS ARMOR3D REP <sup>40,41</sup>
Chlorophyll	SOM-FFN	<a href="https://www.globcolour.info/">https://www.globcolour.info/</a>	Globcolour <sup>58,59</sup>
Marine boundary layer atmospheric xCO <sub>2</sub>	SOM-FFN and air-sea flux	<a href="ftp://aftp.cmdl.noaa.gov/data/trace_gases/co2/flask/surface/">ftp://aftp.cmdl.noaa.gov/data/trace_gases/co2/flask/surface/</a>	NOAA ESRL <sup>50,51</sup>
Sea ice concentration	Air-sea flux and pseudo observations	<a href="https://nsidc.org/data/g02202/versions/3">https://nsidc.org/data/g02202/versions/3</a>	NOAA/NSIDC v3 <sup>60,61</sup>
Sea level atmospheric pressure	Air-sea flux	<a href="http://www.esrl.noaa.gov/psd/data/gridded/data.ncep_reanalysis.surface.html">www.esrl.noaa.gov/psd/data/gridded/data.ncep_reanalysis.surface.html</a>	NCEP reanalysis <sup>53</sup>
Winds	Air-sea flux	<a href="https://www.remss.com/measurements/ccmp/">https://www.remss.com/measurements/ccmp/</a>	CCMP <sup>46,47</sup>
Temperature and salinity	Pseudo observations	<a href="https://sio-argo.ucsd.edu/RG_Climatology.html">https://sio-argo.ucsd.edu/RG_Climatology.html</a>	Roemmich and Gilson climatology <sup>34</sup>
Biogeochemical tracers	Pseudo observations	<a href="https://www.glodap.info/">https://www.glodap.info/</a>	GLODAP v2.2021 <sup>14,18</sup>
Oxygen	Pseudo observations adjustment for surface oxygen undersaturation	<a href="https://soccompu.princeton.edu/www/index.html">https://soccompu.princeton.edu/www/index.html</a>	SOCCOM float data <sup>11,32</sup>

CO<sub>2</sub> concentration), and perhaps least significantly, the additional pCO<sub>2</sub> and pseudo-data used here.

One potential caveat introduced by the pseudo observations (discussed in detail in previous work<sup>13</sup>), is the possibility of a bias in the estimated air-sea CO<sub>2</sub> fluxes. We therefore explored the possibility of bias in the pseudo observations by validating them with a data-assimilating biogeochemical ocean model (see Methods section), and concluded it was neglectable, confirming an earlier validation with direct winter observations<sup>13</sup>.

We have found that the incorporation of pseudo observations that increase winter data coverage at high latitudes from 2004 onwards does not significantly alter the Southern Ocean carbon sink variability over the reconstructed period, suggesting it is relatively well constrained by sparse and seasonally-biased observations using this method. The pseudo observations influence our estimates of the air-sea flux pre-2004 by modifying relationships between fCO<sub>2</sub> and its driver variables in the SOM-FFN reconstruction, but their effect diminishes with time and disappears by the year 2001 (Figs. 3b and 5b). Our conclusions about the robustness of multiyear variability to additional data coverage are therefore limited to the most recent period encompassing the reinvigoration and subsequent stagnation of the sink. A definitively constrained estimate of the variability in the most data-sparse period before 2000 remains elusive.

While introducing the additional observational constraints lowers the air-sea CO<sub>2</sub> uptake in the Southern Ocean, this reduction is more than compensated by corrections made to account for mass boundary layer gradients in temperature and salinity. Those increased the estimated Southern Ocean carbon sink south of 35°S by ~0.3 PgC yr<sup>-1</sup>, balancing the lower uptake induced by the wintertime observations and bringing it in line with atmospheric inversions. Our results confirm earlier conclusions<sup>21</sup> suggesting these corrections are important for understanding the ocean carbon sink.

## Methods

**Calculation and validation of pseudo observations.** The construction of pseudo observations of wintertime surface pCO<sub>2</sub> is described in detail in section 2.1 of an earlier study<sup>13</sup>; here we provide a summary of the method and any differences from that paper. Using data from the Global Ocean Data Analysis Project (GLODAP)<sup>18</sup>,

we identify a Temperature Minimum Layer in summertime temperature profiles south of the Antarctic Polar Front (APF). These are treated as having the preserved properties of the surface waters at the same latitude and longitude in the previous September. We then adjust the concentration of dissolved inorganic carbon (DIC) from the Temperature Minimum Layer for biological activity that occurred in the water mass since last winter using the apparent oxygen utilisation (AOU) and a Redfield ratio RR<sub>C:O</sub>, obtaining an estimate for the wintertime surface DIC concentration. We also apply two corrections to the AOU value to account for oxygen undersaturation when the Temperature Minimum Layer water was last in contact with the atmosphere. The first is a uniform correction of -13.5 μmol kg<sup>-1</sup>, applied away from sea ice, that derives from the mean oxygen undersaturation, determined from SOCCOM float data<sup>11,32</sup>, in the top 10 m in the region between the APF and the sea ice. The second correction, applied in sea ice, takes the form  $AOU_{corr} = AOU(1 - aC_{ice}^b)$ , in which  $C_{ice}$  is the sea ice concentration, and  $a$  and  $b$  are parameters we will optimise during the validation phase (see next paragraph). We then combine the extrapolated wintertime surface DIC estimate with an estimate of total alkalinity derived from a locally interpolated alkalinity regression for global alkalinity estimation (LIAR)<sup>33</sup>, and with surface temperature and salinity from a gridded Argo product<sup>34</sup>, and other tracers from GLODAP, to calculate pCO<sub>2</sub> using MATLAB CO<sub>2</sub>SYS software<sup>35</sup>. The Argo product used in the construction of the pseudo observations spans the years 2004–2018, and consequently this period is the focus of our study.

In order to establish optimal parameters  $a$  and  $b$  for the AOU correction in sea ice, and to further validate the method of calculating pseudo observations beyond what was carried out previously<sup>13</sup>, we use model output from a data-assimilating ocean state estimate with biogeochemistry, B-SOSE<sup>19</sup>. We use the model output to create pseudo observations of DIC in exactly the same manner as had been done using observations, except that we produce model pseudo observations covering the entire region south of the APF for the years 2008–2011 (the limited time range is for convenience, reflecting the packaging of downloaded data; in practice the large number of available model points for validation means this is unlikely to be a limitation). This produces 6791 pseudo observations, which we compare with their wintertime equivalents in the previous September, at the same latitude/longitude grid cell at the surface. We then repeated the calculation for a range of parameters  $a$ , and  $b$ , and two alternative values of RR<sub>C:O</sub>, each time calculating the resulting RMSE between the set of model pseudo observations and their wintertime equivalents. The optimised parameters are  $a = 0.8$ ,  $b = 0.1$ , and RR<sub>C:O</sub> = 106:–138, giving an RMSE of 11.4 μmol kg<sup>-1</sup> and a bias of 0.49 μmol kg<sup>-1</sup>. This compares to  $a = 0.9$ ,  $b = 0.3$  and the same RR<sub>C:O</sub> giving an RMSE of 13.9 μmol kg<sup>-1</sup> and bias of 4.1 μmol kg<sup>-1</sup> for the validation carried out in the earlier paper using a small number of real world observations<sup>13</sup>. In addition, we tested the effect of fixing RR<sub>C:O</sub> to 106:–150, equivalent to the fixed values used by B-SOSE. This gives optimised values of  $a = 0.8$  and  $b = 0.2$ , and the resulting set of pseudo observations produces results in terms of the mean seasonal cycle and annual mean trends of pCO<sub>2</sub> that are indistinguishable from those we have presented on Fig. 2.

Having established  $a$ ,  $b$  and RR<sub>C:O</sub>, we then run a 100-member MonteCarlo ensemble to create pseudo observations of surface pCO<sub>2</sub> while propagating uncertainties in temperature, salinity, AOU and other parameters described previously<sup>13</sup> through the calculation. Next, we take a mean of the ensemble to

produce a set of 798 pseudo observations of wintertime surface  $p\text{CO}_2$  (increased from 760 in the earlier study). We also calculate additional surface  $p\text{CO}_2$  values from wintertime surface GLODAP observations, as in the earlier study<sup>13</sup>. We then convert the  $p\text{CO}_2$  values to  $f\text{CO}_2$  using the formula

$$f\text{CO}_2 = p\text{CO}_2 \exp\left[\frac{B \times 10^{-6} + 2\delta \times 10^{-6} P}{R T}\right] \quad (1)$$

where  $B = -1636.75 + 12.0408 T - 3.16528 \times 10^{-5} T^2 + 3.16528 \times 10^{-5} T^3$ ,  $T$  is the wintertime surface temperature from the Argo product associated with the pseudo observation,  $\delta = 57.7 - 0.118 T$ ,  $P = 1013.25$  is the atmospheric pressure in millibar, and  $R = 8.31447$ . Finally, we combine the GLODAP-derived and pseudo observations of  $f\text{CO}_2$  with the SOCAT  $1^\circ \times 1^\circ$  monthly gridded product by taking the mean of any GLODAP-derived or pseudo observations within a grid cell and the SOCAT  $f\text{CO}_2$  value for that grid cell, if there is one, for September of the year to which those pseudo observations correspond.

**Surface  $f\text{CO}_2$  correction for surface temperature gradients.** The SOCAT gridded  $f\text{CO}_2$  have been corrected for a temperature difference between the depth of the observations several metres below the surface and the base of surface mass boundary layer (MBL), as in a recent study<sup>21</sup>. The temperature at the base of the MBL is obtained from the NOAA Optimum Interpolation (OI) product v2<sup>36</sup>. We also correct our GLODAP-derived and pseudo observations of  $p\text{CO}_2$  to the NOAA OI temperatures, following earlier work<sup>37</sup>, using the formula:

$$p\text{CO}_2^{\text{corr}} = p\text{CO}_2 \exp\left[0.0433(T^{\text{OI}} - T) - 4.35 \times 10^{-5}(T^{\text{OI}2} - T^2)\right] \quad (2)$$

where  $T^{\text{OI}}$  is the NOAA OI product temperature at the latitude, longitude and month corresponding to the  $p\text{CO}_2$  value.

**Gap-filling and mapping surface  $f\text{CO}_2$ .** In the original pseudo observations study<sup>13</sup>, a multiple linear regression (MLR) was used to gap-fill a combination of SOCAT, GLODAP-derived and pseudo observations of surface  $p\text{CO}_2$  and then map them for the Southern Ocean in order to calculate air-sea  $\text{CO}_2$  fluxes<sup>13</sup>. This allowed for MonteCarlo ensembles of the fluxes to be conducted to explore the uncertainties due to the pseudo observation calculation in particular, since the MLR is comparatively computationally inexpensive. However, the MLR is limited in its ability to fit the  $\text{CO}_2$  data, which in general has nonlinear relationships to its driver variables, and in particular was unable to reproduce the interannual variability captured by other studies. Here we employ a Self-Organising-Map Feed-Forward-Neural-network (SOM-FFN) method, applied to a combination of SOCAT, GLODAP-derived and pseudo  $f\text{CO}_2$  values constructed as described earlier in this text, to produce time-varying maps of surface  $f\text{CO}_2$  for the Southern Ocean south of  $35^\circ\text{S}$  from 1993–2018 from which to calculate air-sea  $\text{CO}_2$  fluxes. The SOM-FFN method has previously been described in detail<sup>7,38,39</sup>; here we will detail only the differences with this work.

The time period of 1993–2018 for which we have run the SOM-FFN is motivated by the use of a time-varying mixed layer depth product that starts in 1993<sup>40,41</sup>, and the inclusion of the pseudo observations that have been produced for the period 2004–2018. While we present results for 1993–2018 we focus our analysis on the later period covered by the pseudo observations. The use of interannually varying mixed layer depths is an important distinction from previous work<sup>39</sup> which used a climatology; other datasets that we use to drive the SOM-FFN are summarised in Table 1. Note also that we use the NOAA OI sea surface temperature (SST) product for the boundary layer temperature correction, for driving the SOM-FFN, and for calculating the air-sea flux (see next subsection).

We carry out ensembles of SOM-FFN runs with four members each to explore the sensitivity of the method to certain aspects of the setup. We use either only data south of  $35^\circ\text{S}$  as driver variables, or data south of  $10^\circ\text{S}$ . The latter latitude range provides some more data for the SOM-FFN to draw from, particularly benefitting gap-filling in the mid latitude SouthEast Pacific where  $f\text{CO}_2$  coverage is especially sparse, while excluding some data from near the equator that was found to cause problems for the regressions. We did not include runs using the global datasets because these require significantly more computation time, and results may be unduly influenced by data from regions remote to the Southern Ocean. We also test two methods of organising the data into biogeochemical provinces (known as biomes) in the first stage of the SOM-FFN. The first uses a Self Organising Map (the ‘SOM’ part) employed in earlier studies<sup>7,39,42</sup>, which is time-varying, and the second uses previously published biomes<sup>43,44</sup> which are fixed in time (see Supplementary Fig. S7). We then take an ensemble mean of the mapped  $f\text{CO}_2$  from the 4 combinations of these configurations (south of  $35^\circ\text{S}$  vs south of  $10^\circ\text{S}$  data and fixed vs SOM biomes) to calculate air-sea  $\text{CO}_2$  fluxes. There is a spread in the flux estimates produced by the individual ensemble members, but it is significantly smaller than the flux uncertainties calculated as described in the next section, and it is not included in our results. We then repeat the ensembles with two further variations on our setup: one with and one without the inclusion of the pseudo observations to drive the SOM-FFN mapping. Another final distinction between the SOM-FFN setup in our study and earlier work<sup>7,38,39</sup> is that we do not use the  $p\text{CO}_2$  climatology of Takahashi<sup>45</sup> as a training variable for the SOM, so that our results remain independent of the gap-filling used to produce that climatology. The

RMSE between the mapped SOM-FFN  $f\text{CO}_2$  and the equivalent values from the gridded SOCAT product at the same latitude, longitude and month, for the 4 runs including pseudo observations ranges from 17.63  $\mu\text{atm}$  to 18.77  $\mu\text{atm}$ . In general, the RMSE for the region worsened in tests where the Takahashi data were used in the SOM phase.

**Calculation of air-sea fluxes and uncertainties.** Following a recent study<sup>21</sup>, we use the following equation to calculate the air-sea flux of  $\text{CO}_2$  (values are positive from sea to air):

$$F_{\text{CO}_2} = (1 - C_{ice})k(C_{\text{SW}} - C_{\text{ATM}}) \quad (3)$$

where  $k$  is the gas transfer velocity, and  $C_{\text{SW}}$  and  $C_{\text{ATM}}$  are the concentration of dissolved  $\text{CO}_2$  at the base of the MBL and the air-sea interface, respectively. The gas transfer velocity is a function of wind speed from the CCMP wind product<sup>46–48</sup> and SST from NOAA OI v2<sup>36</sup>, and is calculated according to an established method<sup>10</sup>. The CCMP winds are at  $0.25^\circ$  and 6-hourly resolution, and are interpolated onto a  $1^\circ$  grid to match the other inputs to the gas transfer calculation. The temporal resolution is limited by the satellites, but 6-hourly wind speeds are squared before being monthly averaged.  $C_{\text{SW}}$  is calculated as  $\alpha_{\text{SW}}f\text{CO}_{2-\text{SW}}$ , where  $\alpha_{\text{SW}}$  is the solubility of  $\text{CO}_2$  at the temperature and salinity at the base of the MBL, following an established method<sup>49</sup> and using the data products outlined in Table 1, and  $f\text{CO}_{2-\text{SW}}$  is the  $f\text{CO}_2$  value also at the base of the MBL. Similarly,  $C_{\text{ATM}}$  is calculated as  $\alpha_{\text{AS}}f\text{CO}_{2-\text{ATM}}$ , where in this case  $\alpha_{\text{AS}}$  is the solubility at the air-sea interface, and  $f\text{CO}_{2-\text{ATM}}$  is the atmospheric  $f\text{CO}_2$ , which has been calculated from the NOAA ESRL xCO<sub>2</sub> product<sup>50,51</sup> according to an established method<sup>52</sup> and using NOAA OI SST<sup>36</sup> and NCEP sea level atmospheric pressure<sup>53</sup>. We also correct the surface temperature and salinity used to calculate  $\alpha_{\text{AS}}$  and  $k$  by  $-0.17^\circ\text{C}$  and 0.1 PSU, to approximate the conditions at the air-sea interface.

The three main sources of uncertainty in the calculation of air-sea  $\text{CO}_2$  fluxes are: the uncertainty from gridding SOCAT  $f\text{CO}_2$  observations onto the  $1^\circ \times 1^\circ$  grid, the uncertainty from the SOM-FFN mapping, and the uncertainty on the gas transfer velocity. We calculate the gridding error following earlier work on the SOM-FFN method<sup>7</sup>, obtaining a value of  $\text{err}_{\text{grid}} = 0.27 \mu\text{atm}$  for the Southern Ocean. We calculate the mapping error following another SOM-FFN study<sup>54</sup>, obtaining a value of  $\text{err}_{\text{map}} = 2.00 \mu\text{atm}$ . These two uncertainties are combined to give an overall  $f\text{CO}_2$  uncertainty of  $2.02 \mu\text{atm}$ . We then add/subtract this uniform value from the ensemble mean  $f\text{CO}_2$  to get upper and lower bounds on the full  $f\text{CO}_2$  field. We then combine the upper and lower bound  $f\text{CO}_2$  estimates with a central estimate for the gas transfer velocities, producing two estimates of the air-sea flux. We then take the uncertainty on the fluxes due to the  $f\text{CO}_2$  uncertainty as the difference between the upper bound and central estimate fluxes (or the difference between the central and lower bound estimates; since the errors are normally distributed they are identical). We further combine the central estimate of the  $f\text{CO}_2$  field with upper and lower bounds on  $k$ , assuming 10% uncertainty on  $k$  as suggested by a recent review<sup>55</sup>, producing two more estimates of the air-sea flux. We then take the uncertainty on the fluxes due to the  $k$  uncertainty as the difference between this last upper bound and the central flux estimate. Finally, we combined the relative uncertainties for the flux field due to the  $p\text{CO}_2$  and  $k$  uncertainties to produce the total uncertainty on the flux.

**Atmospheric inversion.** The atmospheric inverse  $\text{CO}_2$  flux estimates used in this analysis follow previously published methods<sup>23</sup>. The methodology employs the GEOS-Chem atmospheric transport model in combination with the Localized Ensemble Transform Kalman Filter (LETKF) data assimilation system and atmospheric  $\text{CO}_2$  observations from the NOAA-ESRL surface network<sup>50</sup>. We use optimized monthly fluxes from the combined GEOSChem-LETKF system at a spatial resolution of  $2^\circ$  latitude by  $2.5^\circ$  longitude. For the comparisons of this study we utilize posterior flux estimates for the Southern Ocean derived from two alternative representations of the prior ocean flux<sup>21,39</sup>. Further details on the GEOSChem-LETKF system are available in the related methods paper<sup>23</sup>.

## Data availability

All of the data used in this study are freely available and downloadable at locations listed in Table 1. The  $p\text{CO}_2$  and air-sea  $\text{CO}_2$  flux product (<https://doi.org/10.25921/zs6b-ng95>) and the pseudo observations (<https://doi.org/10.25921/d44p-fv85>) are available for download at the NCEI OCADS national database <https://www.ncei.noaa.gov/access/ocean-carbon-acidification-data-system/>.

## Code availability

Analysis was carried out in Matlab using standard routines and functions, and scripts used are available on request from the corresponding author.

Received: 11 February 2022; Accepted: 17 October 2022;

Published online: 04 November 2022



## References

- Friedlingstein, P. et al. Global Carbon Budget 2020. *Earth Syst. Sci. Data* **12**, 3269–3340 (2020).
- Devries, T. The oceanic anthropogenic CO<sub>2</sub> sink: storage, air-sea fluxes, and transports over the industrial era. *Global Biogeochem. Cycles* **28**, 631–647 (2014).
- Masson-Delmotte, V. et al. *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* (2021).
- Hauck, J. et al. Consistency and challenges in the ocean carbon sink estimate for the global carbon budget. *Front. Mar. Sci.* **7**, 1–22 (2020).
- Bakker, D. C. E. et al. A multi-decade record of high-quality fCO<sub>2</sub> data in version 3 of the Surface Ocean CO<sub>2</sub> Atlas (SOCAT). *Earth Syst. Sci. Data* **8**, 383–413 (2016).
- Gregor, L. & Gruber, N. 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 (2021).
- Landschützer, P., Gruber, N., Bakker, D. C. E. & Schuster, U. Recent variability of the global ocean carbon sink. *Global Biogeochem. Cycles* **28**, 927–949 (2014).
- Rödenbeck, C. et al. 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 Sci.* **9**, 193–216 (2013).
- Wang, Y. et al. Carbon sinks and variations of pCO<sub>2</sub> in the Southern Ocean from 1998 to 2018 based on a deep learning approach. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens* **14**, 3495–3503 (2021).
- Nightingale, P. D. et al. In situ evaluation of air-sea gas exchange parameterizations using novel conservative and volatile tracers. *Glob. Biogeochem. Cycles* **14**, 373–387 (2000).
- Johnson, K. S. et al. Biogeochemical sensor performance in the SOCCOM profiling float array. *J. Geophys. Res. Oceans* **122**, 6416–6436 (2017).
- Sutton, A. J., Williams, N. L. & Tilbrook, B. Constraining southern ocean CO<sub>2</sub> flux uncertainty using uncrewed surface vehicle observations. *Geophys. Res. Lett.* **48**, 1–9 (2021).
- Mackay, N. & Watson, A. Winter air-sea CO<sub>2</sub> fluxes constructed from summer observations of the polar southern ocean suggest weak outgassing. *J. Geophys. Res. Ocean.* **126**, 1–25 (2021).
- Key, R. M. et al. A global ocean carbon climatology: results from Global Data Analysis Project (GLODAP). *Glob. Biogeochem. Cycles* **18**, 1–23 (2004).
- Olsen, A. et al. GLODAPv2.2019 – an update of GLODAPv2. *Earth Syst. Sci. Data* **11**, 1437–1461 (2019).
- Bushinsky, S. M. et al. Reassessing southern ocean air-sea CO<sub>2</sub> flux estimates with the addition of biogeochemical float observations. *Glob. Biogeochem. Cycles* **33**, 1370–1388 (2019).
- Landschützer, P. et al. A neural network-based estimate of the seasonal to inter-annual. *Biogeosciences* **2**, 7793–7815 (2013).
- Lauvset, S. K. et al. An updated version of the global interior ocean biogeochemical data product, GLODAPv2.2021. *Earth Syst. Sci. Data Discuss.* **2021**, 1–32 (2021).
- Verdy, A. & Mazloff, M. R. A data assimilating model for estimating Southern Ocean biogeochemistry. *J. Geophys. Res. Ocean.* **122**, 6968–6988 (2017).
- Jones, S. D., Le Quéré, C. & Rödenbeck, C. Autocorrelation characteristics of surface ocean pCO<sub>2</sub> and air-sea CO<sub>2</sub> fluxes. *Glob. Biogeochem. Cycles* **26**, GB2042 (2012).
- Watson, A. J. et al. Revised estimates of ocean-atmosphere CO<sub>2</sub> flux are consistent with ocean carbon inventory. *Nat. Commun.* **11**, 1–6 (2020).
- Gruber, N. et al. The oceanic sink for anthropogenic CO<sub>2</sub> from 1994 to 2007. *Science* **363**, 1193–1199 (2019).
- Chen, Z. et al. Variability of North Atlantic CO<sub>2</sub> fluxes for the 2000–2017 period estimated from atmospheric inverse analyses. *Biogeosciences* **18**, 4549–4570 (2021).
- Le Quéré, C. et al. Saturation of the southern ocean CO<sub>2</sub> sink due to recent climate change. *Science* **316**, 1735–1738 (2007).
- Landschützer, P. et al. The reinvigoration of the Southern Ocean carbon sink. *Science* **349**, 1221–1224 (2015).
- Kepler, L. & Landschützer, P. Regional wind variability modulates the southern ocean carbon sink. *Sci. Rep.* **9**, 7384 (2019).
- Gruber, N., Landschützer, P. & Lovenduski, N. S. The variable southern ocean carbon sink. *Ann. Rev. Mar. Sci.* **11**, 159–186 (2019).
- McKinley, G. A., Fay, A. R., Eddebbar, Y. A., Gloege, L. & Lovenduski, N. S. External forcing explains recent decadal variability of the ocean carbon sink. *AGU Adv.* **1**, e2019AV000149 (2020).
- DeVries, T., Holzer, M. & Primeau, F. Recent increase in oceanic carbon uptake driven by weaker upper-ocean overturning. *Nature* **542**, 215–218 (2017).
- Gregor, L., Kok, S. & Monteiro, P. M. S. Interannual drivers of the seasonal cycle of CO<sub>2</sub> in the Southern Ocean. *Biogeosciences* **15**, 2361–2378 (2018).
- Long, M. C. et al. Strong Southern Ocean carbon uptake evident in airborne observations. *Science* **374**, 1275–1280 (2021).
- Johnson, K. S. et al. SOCCOM float data - Snapshot 2017-06-06. In *Southern Ocean Carbon and Climate Observations and Modeling (SOCCOM) Float Data Archive, UC San Diego Library Digital Collections*. <https://doi.org/10.6075/J0348H8K> (2017).
- Carter, B. R., Williams, N. L., Gray, A. R. & Feely, R. A. Locally interpolated alkalinity regression for global alkalinity estimation. *Limnol. Oceanogr. Methods* **14**, 268–277 (2016).
- Roemmich, D. & Gilson, J. The 2004–2008 mean and annual cycle of temperature, salinity, and steric height in the global ocean from the Argo Program. *Prog. Oceanogr.* **82**, 81–100 (2009).
- van Heuven, S., Pierrot, D., Rae, J. W. B., Lewis, E. & Wallace, D. W. R. *MATLAB Program Developed for CO<sub>2</sub> System Calculations. ORNL/CDIAC-105b*. Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, U.S. Department of Energy, Oak Ridge, Tennessee. (2011).
- Reynolds, R. W., Rayner, N. A., Smith, T. M., Stokes, D. C. & Wang, W. An improved in situ and satellite SST analysis for climate. *J. Clim.* **15**, 1609–1625 (2002).
- Goddijn-Murphy, L. M., Woolf, D. K., Land, P. E., Shutler, J. D. & Donlon, C. The OceanFlux Greenhouse Gases methodology for deriving a sea surface climatology of CO<sub>2</sub> fugacity in support of air-sea gas flux studies. *Ocean Sci.* **11**, 519–541 (2015).
- Landschützer, P. et al. A neural network-based estimate of the seasonal to inter-annual variability of the Atlantic Ocean carbon sink. *Biogeosciences* **10**, 7793–7815 (2013).
- Landschützer, P., Gruber, N. & Bakker, D. C. E. Decadal variations and trends of the global ocean carbon sink. *Glob. Biogeochem. Cycles* **30**, 1396–1417 (2016).
- Guinehut, S., Dhomp, A.-L., Larnicol, G. & Le Traon, P.-Y. High resolution 3-D temperature and salinity fields derived from in situ and satellite observations. *Ocean Sci.* **8**, 845–857 (2012).
- Mulet, S., Rio, M.-H., Mignot, A., Guinehut, S. & Morrow, R. A new estimate of the global 3D geostrophic ocean circulation based on satellite data and in situ measurements. *Deep Sea Res. Part II Top. Stud. Oceanogr.* **77–80**, 70–81 (2012).
- Landschützer, P. et al. A neural network-based estimate of the seasonal to inter-annual variability of the Atlantic Ocean carbon sink. *Biogeosciences* **10**, 7793–7815 (2013).
- Fay, A. R. & McKinley, G. A. Global open-ocean biomes: mean and temporal variability. *Earth Syst. Sci. Data* **6**, 273–284 (2014).
- Fay, A. R. & McKinley, G. A. Global Ocean Biomes: mean and time-varying maps (NetCDF 7.8 MB). Supplement to: Fay, AR; McKinley, GA (2014): Global open-ocean biomes: mean and temporal variability. *Earth Syst. Sci. Data* **6**, 273–284 (2014).
- Takahashi, T. et al. Climatological mean and decadal change in surface ocean pCO<sub>2</sub>, and net sea–air CO<sub>2</sub> flux over the global oceans. *Deep Sea Res. Part II Top. Stud. Oceanogr.* **56**, 554–577 (2009).
- Atlas, R. et al. A cross-calibrated, multiplatform ocean surface wind velocity product for meteorological and oceanographic applications. *Bull. Am. Meteorol. Soc.* **92**, 157–174 (2011).
- Mears, C. A. et al. A near-real-time version of the Cross-Calibrated Multiplatform (CCMP) ocean surface wind velocity data set. *J. Geophys. Res. Ocean.* **124**, 6997–7010 (2019).
- Wentz, F. J. et al. *Remote Sensing Systems Cross-Calibrated Multi-Platform (CCMP) 6-hourly Ocean Vector Wind Analysis Product On 0.25 deg Grid, Version 2.0*. Remote Sensing Systems, Santa Rosa, CA. (2015).
- Weiss, R. F. Carbon dioxide in water and seawater: the solubility of a non-ideal gas. *Mar. Chem.* **2**, 203–215 (1974).
- Masarie, K. A., Peters, W., Jacobson, A. R. & Tans, P. P. ObsPack: a framework for the preparation, delivery, and attribution of atmospheric greenhouse gas measurements. *Earth Syst. Sci. Data* **6**, 375–384 (2014).
- Dlugokencky, E., Thoning, K., Land, P. & Tans, P. NOAA Greenhouse Gas Reference From Atmospheric Carbon Dioxide Dry Air Mole Fractions From The NOAA Esrl Carbon Cycle Cooperative Global Air Sampling Network (NOAA, 2017).
- Cooper, D. J., Watson, A. J. & Ling, R. D. Variation of pCO<sub>2</sub> along a North Atlantic shipping route (U.K. to the Caribbean): A year of automated observations. *Mar. Chem.* **60**, 147–164 (1998).
- Kalnay, E. et al. The NCEP/NCAR 40-year reanalysis project. *Bull. Am. Meteorol. Soc.* **77**, 437–471 (1996).
- Landschützer, P., Gruber, N., Bakker, D. C. E., Stemmler, I. & Six, K. D. Strengthening seasonal marine CO<sub>2</sub> variations due to increasing atmospheric CO<sub>2</sub>. *Nat. Clim. Change* **8**, 146–150 (2018).
- Woolf, D. K. et al. Key uncertainties in the recent air-sea flux of CO<sub>2</sub>. *Glob. Biogeochem. Cycles* **33**, 1548–1563 (2019).
- Good, S. A., Martin, M. J. & Rayner, N. A. EN4: quality controlled ocean temperature and salinity profiles and monthly objective analyses with uncertainty estimates. *J. Geophys. Res. Ocean* **118**, 6704–6716 (2013).



57. Rayner, N. A. Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *J. Geophys. Res.* **108**, 4407 (2003).
58. Maritorena, S., D'Andon, O. H. F., Mangin, A. & Siegel, D. A. Merged satellite ocean color data products using a bio-optical model: Characteristics, benefits and issues. *Remote Sens. Environ.* **114**, 1791–1804 (2010).
59. d'Andon, O. H. F. et al. GlobColour - the European Service for Ocean Colour. in *IGARSS 2009* (2009).
60. Meier, W. N., Fettere, M. S., Mallory, R. D. & Stroeve, J. NOAA/NSIDC Climate Data Record of Passive Microwave Sea Ice Concentration, Version 3. <https://doi.org/10.7265/NS9P2ZTG> (2017).
61. Peng, G., Meier, W. N., Scott, D. J. & Savoie, M. H. A long-term and reproducible passive microwave sea ice concentration data record for climate studies and monitoring. *Earth Syst. Sci. Data* **5**, 311–318 (2013).
62. Freeman, N. M. & Lovenduski, N. S. Mapping the Antarctic Polar Front: weekly realizations from 2002 to 2014. *Earth Syst. Sci. Data* **8**, 191–198 (2016).

## Acknowledgements

The authors acknowledge funding from the U.K. Natural Environment Research Council under the SONATA grant (NE/P021298/1) and the UNICORNS grant (NE/W001543/1). Computational resources for the SOSE were provided by NSF XSEDE resource grant OCE130007. This study has been conducted using E.U. Copernicus Marine Service Information. SOCCOM float data were collected and made freely available by the Southern Ocean Carbon and Climate Observations and Modeling (SOCCOM) Project funded by the National Science Foundation, Division of Polar Programs (NSF PLR – 1425989, extension NSF OPP-1936222), supplemented by NASA, and by the International Argo Program and the NOAA programs that contribute to it (<http://www.argo.ucsd.edu>, <http://argo.jcommops.org>). Argo data were collected and made freely available by the International Argo Program and the national programs that contribute to it (<http://www.argo.ucsd.edu>, <http://argo.jcommops.org>). The Argo Program is part of the Global Ocean Observing System. Argo float data and metadata from Global Data Assembly Centre (Argo GDAC) SEANOE <https://doi.org/10.17882/42182>. CCMP Version-2.0 vector wind analyses are produced by Remote Sensing Systems. Data are available at [www.remss.com](http://www.remss.com). We would also like to thank the reviewers for their thoughtful comments and suggestions that have helped significantly improve the paper.

## Author contributions

N.M.: developed of the pseudo observations method and its application to the SOM-FFN, wrote first draft and later revisions of manuscript. A.J.W.: contributed to method

development and manuscript revision; provided funding for the work. P.S.: supervised the atmospheric inversions and contributed to manuscript. Z.C.: carried out atmospheric inversion work. P.L.: originator of the SOM-FFN method, advised on methods and contributed to manuscript revision.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s43247-022-00592-6>.

**Correspondence** and requests for materials should be addressed to Neill Mackay.

**Peer review information** *Communications Earth & Environment* thanks the anonymous reviewers for their contribution to the peer review of this work. Primary Handling Editors: Olivier Sulpis. Peer reviewer reports are available

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022