https://doi.org/10.5194/bg-2021-344
Preprint. Discussion started: 3 January 2022

Biogeosciences Discussions

The sensitivity of pCO_2 reconstructions in the Southern Ocean to sampling scales: a semi-idealized model sampling and reconstruction approach

Laique M. Djeutchouang[1,2], Nicolette Chang[1], Luke Gregor[3], Marcello Vichi[2], Pedro M. S. Monteiro[1]

[1]SOCCO, CSIR, Rosebank, Cape Town, 7700, South Africa
[2]MARIS, Department of Oceanography, University of Cape Town, Cape Town, 7701, South Africa
[3]Environmental Physics, Institute of Biogeochemistry and Pollutant Dynamics, ETH Zürich, Zürich, 8092, Switzerland

Correspondence to: Laique M. Djeutchouang (merlindjeutchouang@gmail.com)

Abstract. The Southern Ocean is a complex system yet is sparsely sampled in both space and time. These factors raise questions about the confidence in present sampling strategies and associated machine learning (ML) reconstructions. Previous studies have not yielded a clear understanding of the origin of uncertainties and biases for the reconstructions of the partial pressure of carbon dioxide (pCO_2) at the surface ocean (pCO_2^{ocean}). Here, we examine these questions by investigating the sensitivity of pCO_2^{ocean} reconstruction uncertainties and biases to a series of semi-idealized observing system simulation experiments (OSSEs) that simulate spatio-temporal sampling scales of surface ocean pCO_2 in ways that are comparable to ocean CO_2 observing platforms (Ship, Waveglider, Carbon-float, Saildrone). These experiments sampled a high spatial resolution (±10km) coupled physical and biogeochemical model (NEMO-PISCES) within a sub-domain representative of the Sub-Antarctic and Polar Frontal Zones in the Southern Ocean. The reconstructions were done using a two-member ensemble approach that consisted of two machine learning (ML) methods, (1) the feed-forward neural network and (2) the gradient boosting machines. With the baseline observations being from the simulated ships mimicking observations from the Surface Ocean CO_2 Atlas (SOCAT), we applied to each of the scale-sampling simulation scenarios the two-member ensemble method ML2, to reconstruct the full sub-domain pCO_2^{ocean} and assess the reconstruction skill through a statistical comparison of reconstructed pCO_2^{ocean} and model domain mean. The analysis shows that uncertainties and biases for pCO_2^{ocean} reconstructions are very sensitive to both the spatial and temporal scales of pCO_2 sampling in the model domain. The four key findings from our investigation are the following: (1) improving ML-based pCO_2 reconstructions in the Southern Ocean requires simultaneous high resolution observations of the meridional and the seasonal cycle (< 3 days) of pCO_2^{ocean}; (2) Saildrones stand out as the optimal platforms to simultaneously address these requirements; (3) Wavegliders with hourly/daily resolution in pseudo-mooring mode improve on Carbon-floats (10-day period), which suggests that sampling aliases from the low temporal frequency have a greater negative impact on their uncertainties, biases and reconstruction means; and (4) the present summer seasonal sampling biases in SOCAT data in the Southern Ocean may be behind a significant winter bias in the reconstructed seasonal cycle of pCO_2^{ocean}.

https://doi.org/10.5194/bg-2021-344
Preprint. Discussion started: 3 January 2022


Biogeosciences Discussions Open Access EGU

1 Introduction

Since the early 2000's, the Southern Ocean (SO) carbon sink has undergone a reinvigoration characterized by a substantial strengthening as reported by Landschützer et al. (2015), following a decade (the 1990s) of weakening trends (Le Quéré et al., 2007; Canadell et al., 2021). Based on these findings, many studies have been conducted recently to investigate both the mechanisms driving these decadal changes in the SO carbon sink and assess the uncertainties of those estimates (Landschützer et al., 2016; DeVries et al., 2017; Fay et al., 2018; Gregor et al., 2018, 2019; Bushinsky et al., 2019; McKinley et al., 2020). Surface ocean CO_2 observations underlying these variability estimates are very sparse, especially during the wintertime, requiring a substantial number of extrapolations to map and subsequently fill the gaps due to data spareness (Landschützer et al., 2014; Gregor et al., 2017, 2019).

Many empirical approaches such as statistical interpolations and regression methods (Rödenbeck et al., 2014; Iida et al., 2015; Jones et al., 2015) gained attention as alternative methods to ocean biogeochemical models (Lenton et al., 2013) until recently where Machine Learning (ML) approaches have been used increasingly as an alternative (Landschützer et al., 2013, 2014, 2016; Gregor et al., 2017, 2019; Denvil-Sommer et al., 2019). These novel mapping methods all seek to fill the spatial and temporal sampling gaps from existing ship-based surface ocean CO_2 observations by extrapolating the CO_2 partial pressure (pCO_2) at the surface ocean (pCO_2^{ocean}) using prognostic proxy variables (such as satellite-observed and re-analysis-based sea surface temperature, sea surface salinity, mixed layer depth, chlorophyll-a, etc.). The feasibility of these extrapolations is justified through the non-linear relationships between surface ocean pCO_2 and the above-mentioned prognostic variables that may drive changes in the surface ocean pCO_2 (Takahashi et al., 1993).

Historically, surface ocean CO_2 observations were primarily from voluntary observing ships that include research and commercial vessels (Bakker et al., 2012; Pfeil et al., 2013). These pCO_2 observations are thus intrinsically biased by the sampling limitations in space and time for the past several decades covering only ~2% of all the monthly 1º observational grid points (Sabine et al., 2013; Bakker et al., 2016). Mainly due to its remoteness and harsh weather especially during winter, it has been increasingly shown that the SO is the ocean region that contributes the most to these uncertainties in the contemporary estimates of mean annual CO_2 uptake (Ritter et al., 2017; Gregor et al., 2019; Bushinsky et al., 2019; Gloege et al., 2021). For instance, sparse observations in largely inaccessible SO areas, particularly during wintertime, have been the biggest barrier in constraining the seasonal and inter-annual variability of regional and global contemporary ocean-atmosphere CO_2 exchange (Monteiro et al., 2015; Rödenbeck et al., 2015; Bakker et al., 2016; Ritter et al., 2017).

Complementary to the increasing effort in the shipboard CO_2 observations through SOCAT initiative, the ongoing development of autonomous ocean observing systems, such as biogeochemical floats and Wavegliders, has started to significantly improve the spatial and temporal coverage of CO_2 samples in recent years (Monteiro et al., 2015; Bakker et al., 2016; Gray et al., 2018;

https://doi.org/10.5194/bg-2021-344
Preprint. Discussion started: 3 January 2022

Bushinsky et al., 2019). Over the last decade, the advent of a range of new autonomous ocean observing platforms has opened doors towards closing the seasonal and intra-seasonal sampling biases created by the high cost of ship operations in the Southern Ocean outside the summer window (Majkut et al., 2014; Monteiro et al., 2015; Williams et al., 2017; Gray et al., 2018; Bushinsky et al., 2019; Sutton et al., 2021).

Thus resolving the mean seasonal cycle through in situ observations not only is a challenging exercise that has followed several avenues from extrapolating findings from the Drake Passage Time-series (DPT) like in Fay et al. (2018), to utilizing measurements from extended deployments of autonomous ocean observing platforms such as Wavegliders (Monteiro et al., 2015), biogeochemical Argo floats (Williams et al., 2017; Gray et al., 2018; Bushinsky et al., 2019), and more recently Saildrones (Sutton et al., 2021). These advances have allowed the density of the Southern Ocean surface CO_2 observing networks to increase. Consequently, the problem of general sparseness in observations and particularly the sampling biases (Monteiro et al., 2015; Gloege et al., 2021) was partially addressed but not resolved by the ocean CO_2 in-situ observations community (Bushinsky et al., 2019; Sutton et al., 2021). For example, under-sampling in winter by ships has been addressed by the 10-day resolution SOCCOM profiling floats and/or Lagrangian platforms that are carried zonally by water currents (Majkut et al., 2014; Williams et al., 2017; Gray et al., 2018; Bushinsky et al., 2019). Williams et al. (2017), and then Gray et al. (2018) reported on persistent differences found with previous pCO_2 estimates when the ship-based sampling is sparse, especially during winter, though a recent study seems to disagree on the persistence of these differences (Bushinsky et al., 2019). Increase in winter sampling would yield a reduction in the uncertainty levels of surface ocean pCO_2 estimates (Gregor et al., 2019; Bushinsky et al., 2019). Notwithstanding these new platforms, sparse and scale sensitive observations in the Southern Ocean continue to be a barrier in constraining the seasonal cycle and inter-annual variability of surface ocean pCO_2 (Monteiro et al., 2015; Rödenbeck et al., 2015; Sutton et al., 2021).

However, we appear to have reached a limit in terms of improving the uncertainty and biases underlying these reconstructions as reported by Gregor et al. (2019). This convergence of performance measures in existing empirical methods led the authors to the rhetorical question “have we hit the wall?” In practice, high-quality in-situ CO_2 observations like those of the annually collected and compiled within the SOCAT database are fundamental to novel machine learning (ML) methods (Sabine et al., 2013; Bakker et al., 2016), despite the reconstructions being limited by spatial and temporal observational gaps and biased sampling (Gregor et al., 2019).

The aim of this study is to investigate the sensitivity of the pCO_2 reconstructions to the spatio-temporal sampling scales of surface ocean CO_2 observing systems. To do that, we conducted a series of semi-idealized model experiments based on an ensemble of two state-of-the-art machine learning techniques (ML2), then a rigorous assessment of experiments was conducted through testing and understanding of ML2 capabilities. We explore the question set by Gregor et al. (2019) that the prediction uncertainty and biases in contemporary pCO_2 reconstructions are now constrained by the sampling scales achievable by the

https://doi.org/10.5194/bg-2021-344
Preprint. Discussion started: 3 January 2022

existing ocean observing platforms. We make proposals towards significantly advancing machine learning reconstructions "beyond the wall".

2 Materials and methods

2.1 Data source

The data used in this study is from a year-long period of high-resolution ocean model simulations. The ocean model is a regional configuration (BIOPERIANT12-CNCRUN05A-S) of the state-of-the-art ocean modelling framework NEMO (Nucleus for European Modelling Ocean) coupled with the biogeochemical model PISCES (Pelagic Interactions Scheme for Carbon and Ecosystem Studies) which simulates the lower tropic level of the marine ecosystem and the biogeochemical cycles of carbon and nutrients (Aumont et al., 2015). More specifically, we used (1/12)º by (1/12)º-daily simulations of a forced NEMO-PISCES regional Southern Ocean model called BIOPERIANT12 (BP12). There are many prognostic variables including two phytoplankton compartments (diatoms and nanophytoplankton) and a description of the carbonate chemistry in the model. However, we focused only on the variables of particular interest for our study; these are the CO_2 partial pressure (pCO_2) at the surface ocean (pCO_2^{ocean}), and its well-known drivers (Takahashi et al., 1993): sea surface temperature (SST), sea surface salinity (SSS), mixed layer depth (MLD), chlorophyll-a (Chl-a). Their characterization is presented with more details in Table S1.

2.2 Data processing and derived variables

In preparation for the training and validation phases of the machine learning (ML) algorithms, some of the input data are transformed for better interpretation. At first this includes the mixed layer depth (MLD) and chlorophyll-a (Chl-a) data that undergo a log10 transformation to return a distribution closer to a normal distribution (Maritorena et al., 2010; Holte et al., 2017). Secondly, it is substantially beneficial to include only the temporal coordinate (time) as a proxy of pCO_2^{ocean} due to the characteristics of our study area (Fig. 1a) as being a single domain with no regional or clustering subsets. Thus, we did not include spatial coordinates (latitude, longitude) in the predictors although Gregor et al. (2017, 2018) and many studies in Rödenbeck et al. (2015). However, it's important to note that coordinate variables do not drive mechanistic changes in pCO_2^{ocean} according to Gregor et al. (2017).

The inclusion of the temporal coordinate as proxy of pCO_2 was done through a variable transformation that aims to preserve the seasonality of the data. More precisely, the preservation of this seasonality was done by transforming the day-of-the-year (j) as in Gregor et al. (2017); that is,

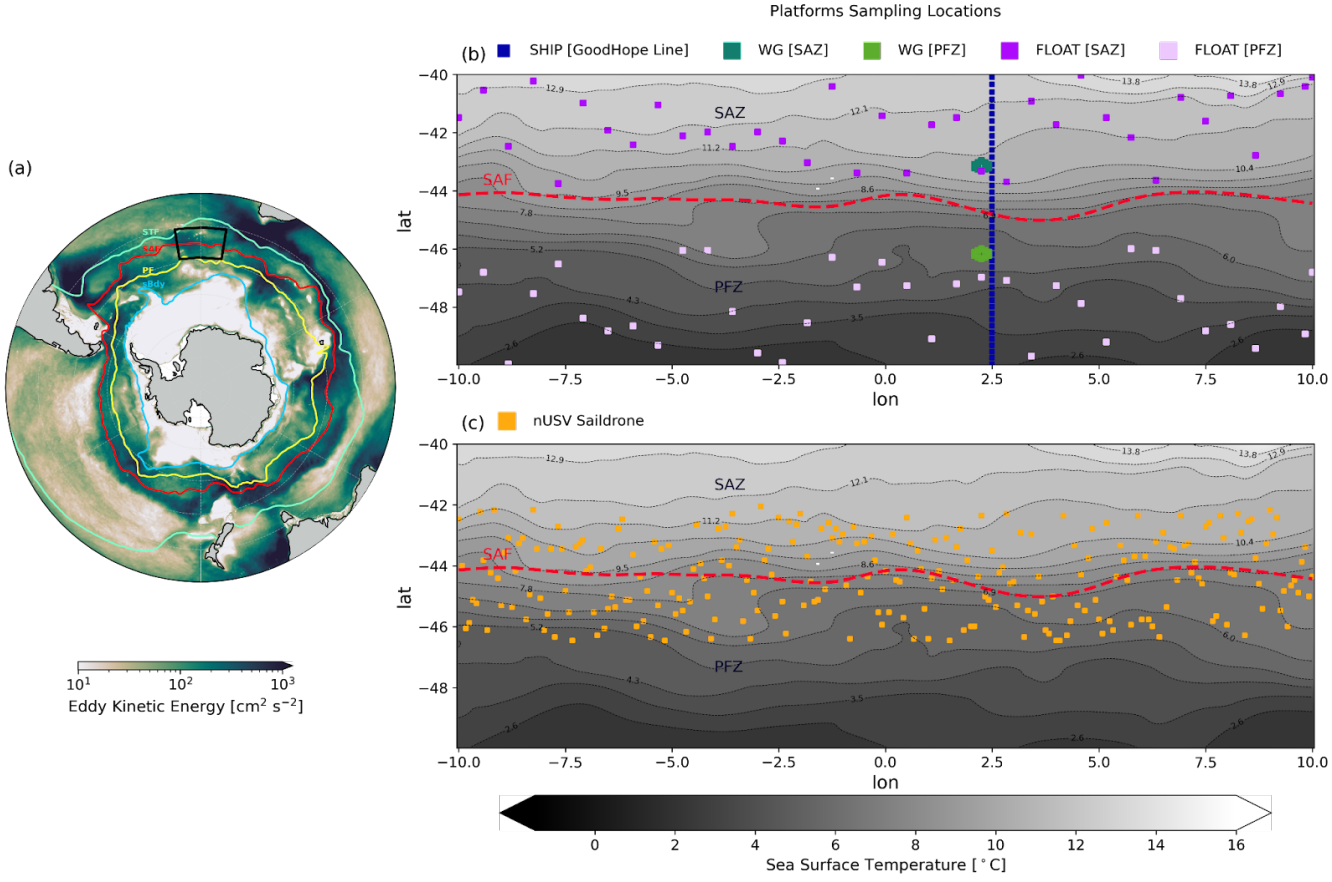
$$J = \left(cos\left(j.\frac{2\pi}{365}\right),\ sin\left(j.\frac{2\pi}{365}\right)\right) \quad (1)$$

https://doi.org/10.5194/bg-2021-344
Preprint. Discussion started: 3 January 2022

130 **2.3 Experimental configurations**

2.3.1 Study region and selection of the experimental domain

The seasonal cycle is known as the strongest mode of natural variability of CO_2, and also the one that most strongly links climate and ocean ecosystems (Mongwe et al., 2018). Given its characteristics that are largely shaped by higher frequency intra-seasonal modes defining the response modes in physics and biogeochemistry components, the Southern Ocean Seasonal
135 Cycle Experiment (SOSCEx) project was created (cf. Sect. S1.1 for more details). As depicted in Fig. S1, the novel aspect of the third phase of SOSCEx was the integration of a multi-platform approach that consisted of combining gliders, ships, floats, satellites, and prognostic models to explore new questions about climate sensitivity of CO_2 and ocean ecosystem dynamics and how these processes are parameterized in forced ocean models such as the NEMO-PISCES, BIOPERIANT12.



140 **Figure 1: Panel (a) is the regional view of the BIOPERIANT12 model simulations with the selected experimental domain (black box) within the annual mean of the Southern Ocean major fronts and the changing conception of the Antarctic Circumpolar Current (ACC), showing the mean annual of eddy kinetic energy (EKE) derived from the model. From the north to south are the mean locations of the named fronts: the Subtropical Front (STF), the Subantarctic Front (SAF), the Polar Front (PF), and the Southern Boundary (SBdy) front (based on** Orsi et al. (1995)**). Colours show the EKE, illustrating the strong steering of the fronts. Panel (b)**

https://doi.org/10.5194/bg-2021-344
Preprint. Discussion started: 3 January 2022


Biogeosciences Discussions

shows the map of the SST in the experimental domain (black box in Fig. 1a) on which are also shown the idealized sampling tracks/locations of the synthetic ocean observing platforms, SHIP, FLOAT and WG as described in the figure legend. Panel (c) shows the sampling tracks of the idealized new unmanned surface vehicle (nUSV) Saildrone within the experimental domain. These locations, marked and coloured according to each corresponding sampling platform, are where we sample the BP12 model data in a way that is comparable to the real-world. The SAF is characterized by the red line (Fig. 1a) and red dashed line (Fig. 1b-c), and it separates the experimental domain into the Sub-Antarctic Zone (SAZ) and Polar Front Zone (PFZ).

This study was designed as a semi-idealized experiment to minimize some of the potential confounding factors on the final estimation of the RMSE, MAE and temporal and spatial biases while evaluating the performance of regression models used to extrapolate surface ocean pCO_2 values. A key part of this design was to remove the normal step of clustering necessary to overcome the spatial and temporal limitations of observations (Fay and McKinley, 2013, 2014; Gregor et al., 2019; Landschützer et al., 2014). Thus, to avoid the clustering step, we chose a domain in the high-resolution (±10km) BIOPERIANT12 forced ocean biogeochemical model that was not only spatially and temporally coherent but also big enough to reflect the spatial and temporal variability necessary to provide sufficient sensitivity to the different sampling strategies. The selected domain, 40ºS - 50ºS and 10ºW - 10ºE (Fig. 1a), is located in the Atlantic sector of the Antarctic Circumpolar Current (ACC) between the Subtropical Front (STF) and the Polar Front (PF) and spans across the Subantarctic Front (SAF) (Fig. 1a). Furthermore, the domain lies within the Sub-Polar Seasonally Stratified (SPSS) biome from Fay and McKinley (2014). The Good Hope repeat hydrography sampling line passes through the domain (Fig. S1) for which sustained annual to bi-annual ship-based observations have been carried out for over a decade, as well as high-resolution carbon glider observations (Monteiro et al., 2015). More specifically, as shown in Fig. 1, our selected domain is divided into the SAZ and the PFZ, inspired from Gray et al. (2018) and Chapman et al. (2020). The SAZ and PFZ, separated by the SAF (red dashed curve in Fig. 1b-c), are respectively referred to as the north and south of the experimental domain.

The oceanographic context of this domain is shown in Fig. 1a, depicting the selected 10º-by-20º domain (black box) in the context of the Southern Ocean major fronts and the eddy kinetic energy (EKE) derived from the BIOPERIANT12 model. This confirms that the domain spans the SAF and is in a region of relatively medium EKE.

2.3.2 Model vs data products: the mean seasonal cycle of pCO_2

One of the major contributions from machine learning methods has been the increased confidence in the constraints for the climatology, interannual variability and trends of pCO_2 in the global ocean and particularly in the Southern Ocean (Hauck et al., 2020; Canadell et al., 2021; Friedlingstein et al., 2021). These semi-independent empirical models converge on the canonical representation of the seasonal cycle of pCO_2 in the Southern Ocean with a maximum positive anomaly for pCO_2 in winter (Jul - Sep), linked to mixed layer deepening and associated entrainment and a maximum negative anomaly in summer linked to the spring - summer net primary production (NPP)(Takahashi et al., 2009; Gregor et al., 2018). The reconstructions

https://doi.org/10.5194/bg-2021-344
Preprint. Discussion started: 3 January 2022

from two such empirical models (Fig. 2b-c) show that the seasonal cycle of pCO_2 in this study's domain is consistent with the broader Southern Ocean (Gregor & Gruber, 2021).

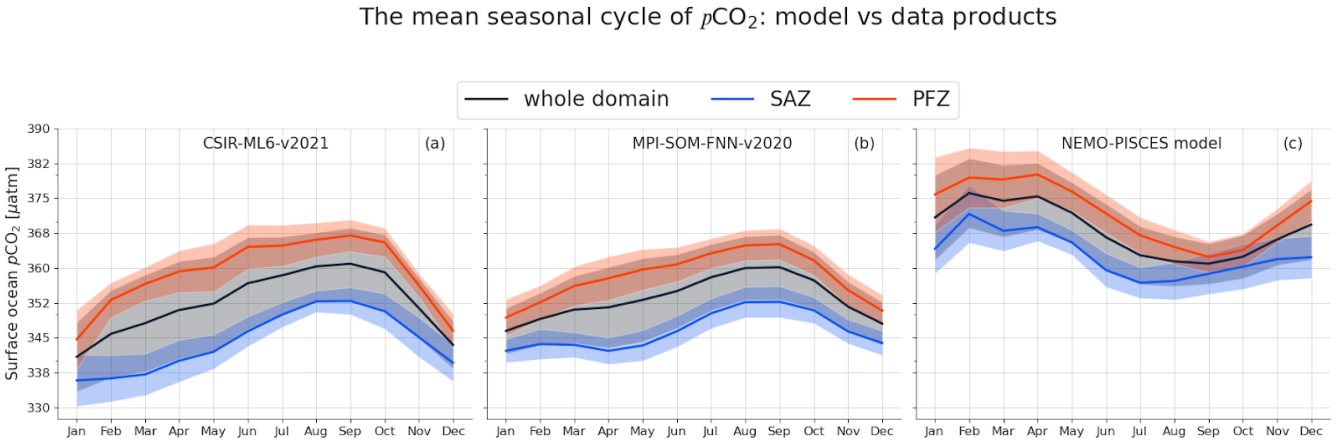


Figure 2: The mean seasonal cycle (SC) for surface ocean pCO_2 from two observation-based products and a high resolution (±10km) forced NEMO-PISCES ocean model (BIOPERIANT12) within the selected experimental domain (Fig. 1a). It contrasts their respective seasonal cycles. Panels (a) and (b) respectively show the mean SCs of the pCO_2 estimates from the two data products: CSIR-ML6-v2021 (Gregor et al., 2019) **and MPI-SOM-FNN-v2020** (Landschützer et al., 2016) **in the whole domain, the SAZ, and the PFZ; and similarly, panel (c) shows the mean SCs of the pCO_2 from the BIOPERIANT12 model.**

Our BP12 model study sub-domain (Fig. 1a) is depicted as a winter-maximum and summer-minimum pCO_2 area by both observational products (Gray et al., 2018; Keppler and Landschützer, 2019) as shown in Fig. 2a-b. This is in sharp contrast with the seasonal cycle climatology from the high-resolution forced ocean model used in this study (Fig. 2c). The basis for this difference is that the high resolution forced ocean model has a seasonal cycle that is largely influenced by the annual cycle of SST (Fig. 2c). This kind of temperature driven model bias for surface ocean pCO_2 is now well recognized in both forced and coupled models in the Southern Ocean (Mongwe et al., 2016, 2018). This study is not concerned with the realistic accuracy of the mean annual or the seasonal cycle of pCO_2 but whether the model represents the range of temporal and spatial modes of variability that can be used to test the sensitivity of the reconstructions to particular choices in the sampling scales. Thus, these phasing differences between models and products are not relevant to the objectives of this study.

2.3.3 Synthetic ocean observing platforms

In designing the sampling scales and strategies we opted to constrain the experiment to realistic and existing observing platforms that can make direct pCO_2^{ocean} or derived (from pH) surface ocean CO_2 observations. More specifically, the existing ocean observing platforms involved in these experiments are the ships, and the following autonomous unmanned surface vehicles (USVs): Carbon-floats, Wavegliders and Saildrones (the new USV), whose simulations are dubbed SHIP, FLOAT, WG, and nUSV, respectively (Fig. 1b-c). The Waveglider is an autonomous USV, developed by Liquid Robotics Inc

https://doi.org/10.5194/bg-2021-344
Preprint. Discussion started: 3 January 2022

(Sunnyvale, California, USA), that is unique in its ability to harness ocean wave and solar energy for platform propulsion (Hine et al., 2009). At sea it operates individually or in fleets delivering real-time data for several months without servicing (Sabine et al., 2020; Grare et al., 2021). Equipped with physical and biogeochemical instruments/sensors, the Waveglider gathers thus ocean data in ways or locations previously either too costly or challenging to operate. Made by Saildrone Inc (Alameda, California, USA), the nUSV Saildrone is an autonomous ocean-going data collection platform navigable via satellite communications and designed for long-range, long duration missions of up to 12 months (Meinig et al., 2016, 2019; Gentemann et al., 2020). It is predominantly powered by wind and solar energy, and equipped with meteorological, ocean physical and biogeochemical sensors for long-range ocean data collection missions (Gentemann et al., 2020) through remote surveying in the toughest of ocean environments such as the Southern Ocean (Meinig et al., 2019; Sutton et al., 2021).

Each of these simulated mentioned sampling platforms had a sampling routing through the domain that closely approximated reality. Ship-based sampling is along a single meridional repeat line (longitude), where repeats could be seasonal and annual (Fig. 1b). Floats followed a zonal sampling distribution that is consistent with the flow of the ACC and a 10-day sampling scale with a limited random meridional mesoscale variability which reflects the eddy kinetic energy (EKE) characteristics of the domain but constrained by the SAF (Fig. 1a-b). Wavegliders were constrained to repeat the pseudo-mooring sampling (25km range) on the ship line (Fig. 1b), which captures the sub-mesoscale gradients but with a high temporal sampling frequency of 1 hour.

2.3.4 Idealized experiment setup

In this paragraph, we briefly describe the experimental scenarios shown in Table S2. We stress again on the fact that these experiments are intentionally made to reproduce the sampling resolutions of their real-world counterparts, not necessarily the spatial resolution in practice but at least the temporal one. We considered the NEMO-PISCES model simulations, BP12, as a realistic representation of the real ocean climate systems within which the pCO_2^{ocean} is known across the entire experimental domain. Based on this, we ask the following question: given measurements of pCO_2^{ocean} as sampled in a real-world scenario by these ocean observing platforms, how sensitive is the sampling distribution and resolution to observation-based estimates of pCO_2^{ocean} at every point across the entire experimental domain?

In these experiments, we simulate the sampling tracks/patterns of the synthetic ocean observing platforms SHIP, FLOAT, WG, and nUSV Saildrone (Fig. 1b-c). We leverage these synthetic sampling systems to sample the BP12 model data inside our selected experimental domain by constraining the experiment to their realistic and existing counterparts. The BP12 model sampled data from each of the sampling scenarios are then used for training and testing of the ML algorithms. The trained ML models are used to reconstruct the pCO_2^{ocean} values of the full experimental domain and compared with original BP12 model field pCO_2^{ocean} in order to assess the anomalies in reconstructed mean annual and seasonal cycles. The idealized ship operates

https://doi.org/10.5194/bg-2021-344
Preprint. Discussion started: 3 January 2022

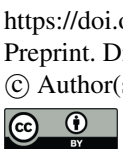
according to the sampling scales and strategies of ships involved in SOCAT collaborative effort. However, here we considered the three following seasonal sampling regimes for the ship platform: (1) summer only, (2) winter and summer, and (3) autumn and spring. Like the real-world scenario, the ship simulation served as our baseline. The idealized carbon-float simulates SOCCOM biogeochemical float with a 10-day sampling cycle. Talley et al. (2019) reported the importance of the water masses and frontal structures in the deployment strategy of autonomous sampling platforms, such as floats, that will likely follow the fronts an eastward trajectory but will seldom cross the front. Therefore, we consider the situation where the idealized floats do not cross the SAF as illustrated in Fig. 1b, even though in reality this might happen due to the occurrence of events such as storms or eddies. We thus considered two deployment and sampling scenarios: (1) in the SAZ, and (2) in the PFZ (Fig. 1b). Given the Lagrangian sampling patterns of an Argo float whose motion is driven by water current, we assume that our idealized float moves eastward and on a trajectory that is a Brownian motion or, more specifically, a random walk (Fig. 1b). The idealized Waveglider operates according to the sampling strategies of the Wavegliders used in the SOSCEx project (cf. Sect. S1.1 for additional details). Like the idealized float, we considered two deployment stations, the first in the SAZ (cf. Fig. 1b, hexagonal patterns in dark green) and the second in the PFZ (cf. Fig. 1b, hexagonal patterns in light green). This idealizes the two deployment scenarios of SOSCEx III gliders (cf. Fig. S1, hexagonal patterns in blue-yellow) that sampled on an hourly basis. However, given the model temporal resolution that is daily, our idealized Wavegliders samples daily. Lastly, we add an idealized Saildrone that simulates the sampling strategies of its real-world counterpart that can sample for up to 12-month ocean data collection missions (Meinig et al., 2019; Gentemann et al., 2020). As with the idealized Waveglider, the Saildrone also samples daily. Further, we assume that leveraging its speed the Saildrone sampling can be done across an ocean front, such as the SAF (Fig. 1c) – a realistic assumption, as the nUSV Saildrones can be piloted remotely. We assumed that all three autonomous sampling platforms sampled year-round in our experimental domain.

In summary, we sample pCO_2^{ocean} and drivers using these above-mentioned synthetic sampling platforms, i.e., SHIP, FLOAT, WG, and nUSV Saildrone (Fig. 1b-c). We emphasize that these experiments are intentionally made to reproduce the sampling resolution of their real-world counterparts, not necessarily their spatial resolution in practice but at least the temporal one. Then we use ML regression techniques to reconstruct the full experimental domain and compare it with the BP12 model truth pCO_2^{ocean} in the full domain to assess the reconstructions as anomalies of mean annual and seasonal cycles, which is a key objective of this work.

2.4 Machine Learning implementation

We use a two-member ensemble method (we call ML2) that consists of two state-of-the-art ML approaches: the Feed-forward Neural Network (FNN) and a variant of Gradient Boosting Decision Tree (GBDT) learning frameworks called Gradient Boosting Machines (GBM). Our choice of FNN method is motivated by its recent success in approximating the surface ocean pCO_2 (Landschützer et al., 2013, 2016; Gregor et al., 2019; Denvil-Sommer et al., 2019). The choice of the GBDT approach

https://doi.org/10.5194/bg-2021-344
Preprint. Discussion started: 3 January 2022



is motivated by its achievement of state-of-the-art performances in many ML tasks (Ke et al., 2017), and also the success of GBDT's previous approaches (Gregor et al., 2019; Gloege et al., 2021; Gregor and Gruber, 2021). We use the Scikit-learn and LightGBM Python packages for our implementation of FNN and GBM, respectively. We thus focus here only on the ensemble average ML2 whose the stacking process is illustrated in Fig. 3a.

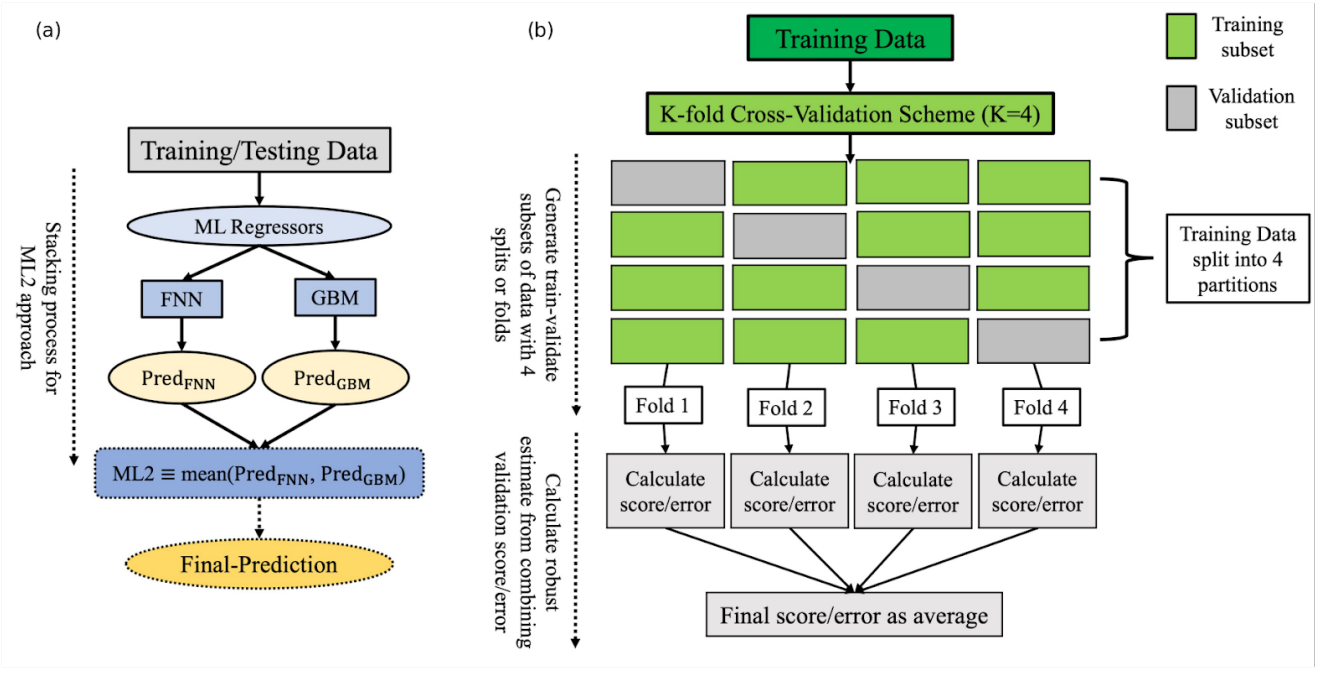


Figure 3: Schematic flow diagram of the two-member ensemble method ML2. Panel (a) shows the schematic representation of the stacking process of the two machine learning (ML) algorithms, FNN and GBM, that make up ML2; panel (b) shows the schematic flow diagram of the K-fold cross-validation (CV) procedure used in hyper-parameter optimization (HOP) of the two members (FNN, GBM) of ML2. To extrapolate from surface ocean pCO_2 samples, ML2 uses full domain coverage model data of the predictor variables: SST, SSS, MLD) and Chl-a. These variables serve as proxies for known processes that affect surface ocean pCO_2 (Takahashi et al., 1993)**.**

Given that the observation size is relatively small, especially for the SHIP experiment, splitting the data into training and testing sets may not capture some key features of the original platform observations. We thus use the entire sampled data for model building instead of splitting the data into two sets. We used a K-fold cross-validation (CV), with K=4, to find the set of hyper-parameters that enabled a better generalization of ML2 (Fig. 3b). Like in Gregor et al. (2019), the CV is applied identically to each of the two-member algorithms (FNN and GBM) except that here, the search for the optimal hyper-parameters is achieved with a Bayes-search CV (BayesSearchCV) instead of a grid-search CV. We make use of the Scikit-optimize Python package for our BayesSearchCV implementation.

https://doi.org/10.5194/bg-2021-344
Preprint. Discussion started: 3 January 2022

2.5 Machine learning regression metrics

Although the choice of the performance measure may seem straightforward and objective, it is often difficult to choose a metric that corresponds well to the desired behaviour of the ML algorithm (Goodfellow et al., 2016). The reconstruction power of the surface ocean pCO_2 of the full experimental domain are thus estimated using a series of four statistical metrics that include the mean bias error (MBE), mean absolute error (MAE), root mean square error (RMSE), and Pearson's correlation coefficient (r) to measure the tendency or strength of estimates and observations to vary together (Stow et al., 2009) or, more technically, to quantify the level at which reconstruction captures the phasing observed in the model truth (Gloege et al., 2021).

The MBE, commonly called bias, is the mean difference between the estimates and the target variable samples. It captures the average bias/error in the predictions and is calculated as follows:

$$MBE = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i), \quad (2)$$

where n is the number of samples, $\hat{y}$ is the model prediction and y is the target variable (in this case, pCO_2^{ocean}).

The MAE denotes the ratio of the L1 norm of the error vector to the number of samples (n). More specifically, the MAE derives from the unaltered magnitude (or absolute value) and provides an estimate of the average magnitude of the error. It is calculated as follows:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|\hat{y}_i - y_i| \quad (3)$$

The RMSE, one of the most popularly used metrics in the climatic and environmental sciences community when dealing with regression modelling problems, is also a measure of the difference between the estimates $\hat{y}_i$ and the target variable samples y_i. It provides an estimate of the variability in the predictions in terms of the fitness with the observed data, and is defined as follows:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2} = \sqrt{MSE} \quad (4)$$

Where MSE is simply the mean square error. For squaring individual errors $e_i = \hat{y}_i - y_i$ $(i = 1, \ldots, n)$, the stated rationale is usually to "disconnect the sign" of e_i so that the magnitudes of errors influence the average error, MSE.

In order to quantify the strength of the linear association between the pCO_2^{ocean} estimates (i.e., $\hat{y}_i$ for $i = 1, \ldots, n$) and observations/known truth (i.e., y_i for $i = 1, \ldots, n$), the Pearson's correlation coefficient (r) is used. Its computing is formulated as follows:

$$r = \frac{1}{(n-1)\sigma_y\sigma_{\hat{y}}}\sum_{i=1}^{n}(y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}}), \quad (5)$$

https://doi.org/10.5194/bg-2021-344
Preprint. Discussion started: 3 January 2022

where σ_y and $\sigma_{\hat{y}}$ are the standard deviations of y and $\hat{y}$, respectively; $\bar{y}$ and $\bar{\hat{y}}$ the means of y and $\hat{y}$, respectively. The correlation coefficient always takes values between -1 and 1, with lower (near -1) and higher (near 1) values of r respectively indicative of how much reconstruction and model are in or out of phase. Values of r that are close to 0 are indicative of no association between the two signals. Therefore, the ideal value for r will be close to one.

325

2.6 Uncertainty decomposition/breakdown

A firm understanding of the uncertainties is required for the purpose of our analysis given that in our study we are dealing with the uncertainties that we cannot fully quantify now as this is on unseen or out-of-sample data like in Gloege et al. (2021). Therefore, it is necessary to distinguish the different types of uncertainties. We assume that our sampled observations are
330 unbiased, and hence the training data sets for surface ocean pCO_2 are considered such as known; and this can be justified by the fact that we have access to the whole data. The terms error and uncertainty are interchangeably used although here the latter is used as an estimate quantifiable against a known value whereas the former characterizes a range of values within which the true value is asserted to lie with some level of confidence (Gregor and Gruber, 2021).

335 The total pCO_2 uncertainty (E) is dealt with as in Gregor and Gruber (2021) where the authors broke down the pCO_2 uncertainties into three key components according to three identified sources of errors associated with the surface ocean carbonate system: (1) the prediction (P), (2) representation (R), and (2) measurement (M) uncertainties. Under the assumption that these components are independent of each other in the pCO_2 error space, E can thus equivalently be expressed as the norm of the vector whose coordinates are P, M and R; that is, the square root of the sum of the squares of these components: $E = \sqrt{P^2 + M^2 + R^2}$. We can remove the contribution of the measurement uncertainty from this equation since we are sampling from a synthetic dataset. Further, we address the representation uncertainty by sampling at a higher resolution (Gregor and Gruber, 2021). Given that we are predicting at high resolution (1/12º daily), the sampling distribution bias due to capturing of large-scale gradients is assumed to be small since we are within the 2-day threshold set by Monteiro et al. (2015). Lastly, we assume that the ML models are the best possible predictors for the given training datasets, since each ML model was trained
345 using best practices(i.e., low in-sample errors calculated from all the training points as shown in the supplementary materials). Therefore, reported RMSEs will be the uncertainties due to sampling bias.

3 Results

In the next sections, the results for the following four sets of semi-idealized model experiments combinations, SHIP, SHIP +
350 FLOAT, SHIP + WG, and SHIP + nUSV are presented in terms of spatial and seasonal cycle anomalies of the annual mean pCO_2 estimates.

https://doi.org/10.5194/bg-2021-344
Preprint. Discussion started: 3 January 2022

3.1 Annual mean seasonal cycle for the domain

The annual mean map for pCO_2 (mean 368.15 µatm; standard deviation 50.5 µatm) shows that the domain is characterized by both meridional and mesoscale variability expected from the mesoscale resolving BIOPERIANT12 model (Fig. 4a). The meridionally distinct SAZ (north of the domain) (< 368.15 µatm) and PFZ (south of the domain) (> 368 µatm), are separated by the Subantarctic front (SAF) (Figs. 1 and 4a). This mean map also highlights the importance of mesoscale gradients in both the SAZ and the PFZ domains (Fig. 4a). The mean seasonal cycles of pCO_2 for the whole domain as well as for the SAZ (lower - blue) and PFZ (higher - red) are depicted in Fig. 4b. It shows that the seasonal cycle of pCO_2^{ocean} is dominated by the influence of the annual cycle of the sea surface temperature (SST) on CO_2 solubility (Munro et al., 2015; Mongwe et al., 2016) with warm late summer (Feb-Apr) and cool late winter (Jul-Sep) (Fig. 4b). The three seasonal cycles (whole domain, SAZ, and PFZ) show coherence in the seasonal amplitude and phasing except that the warming transition from winter to spring occurs two months earlier (Jul) in the SAZ relative to the PFZ (Fig. 4b).

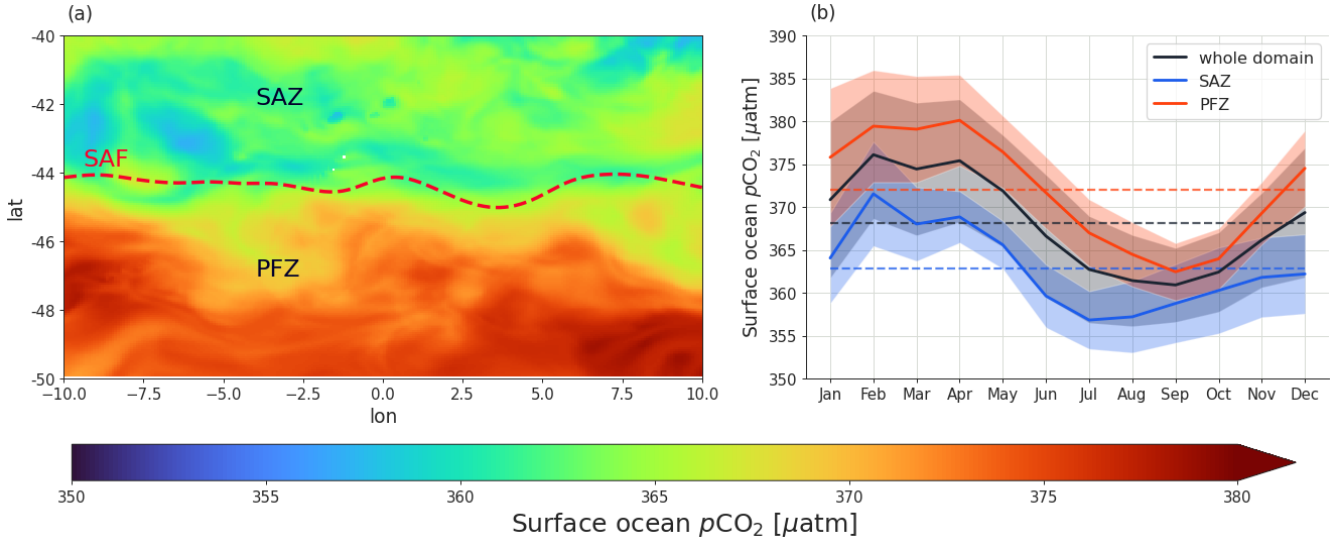


Figure 4: Characterisation of the spatial and temporal surface ocean pCO_2 annual mean state within the selected 10º-by-20º experimental domain located in the northern ACC that corresponds to SPSS biome (Fay and McKinley, 2014) as shown in Fig. 1a. Panel (a) shows the map of mean annual pCO_2 from the BIOPERIANT12 (BP12) model. It shows that the domain is characterized by a regional meridional gradient including the Sub-Antarctic Front (SAF) (red dashed line) as well as mesoscale gradients in both SAZ and PFZ; panel (b) shows the mean seasonal cycles for surface ocean pCO_2 in the BP12 model domains (SAZ, SAF and PFZ) where the dashed lines indicate the magnitude of the annual mean for each domain - 368.16 µatm (domain), 362.85 µatm (SAZ), and 371.78 µatm (PFZ).

Notwithstanding the phasing differences, we still find a comparable winter reconstruction bias in this study (Figs. 2c and 4b) and observation-based products (Fig. 2a-b). Thus, the question is: is the magnitude of the reconstructed winter pCO_2 maximum

https://doi.org/10.5194/bg-2021-344
Preprint. Discussion started: 3 January 2022

realistic or a result of the way the machine learning methods process the summer sampling bias in a system characterized by strong seasonal and intra-seasonal modes of variability?

3.2 Reconstructed mean annual spatial and seasonal cycle anomalies

In order to investigate the anomalies in the reconstruction of the mean annual and seasonal cycles, which are a key objective of this study, we first characterized the anomaly by the mean bias error (MBE) and calculated the MBE at each grid point of the spatial domain. Secondly, we also calculated the anomaly of the seasonal cycle reconstruction in each of the sub-domains. More specifically, we used the seasonal cycle residuals to explore how a systematic anomaly could influence the reconstruction of pCO_2 values at the surface ocean. We performed this calculation for each experiment and their respective reconstructions and also examined their spatial variability.

3.2.1 Semi-idealized SHIP-only observations experiment results

The semi-idealized SHIP-only sampling experiments mimic the largely ship-based SOCAT gridded product to evaluate the sensitivity of the reconstruction uncertainties (RMSE, MAE, MBE/bias) to seasonal meridional sampling scenarios. In each of these scenarios the ship makes two meridional crossings in opposite directions one month apart (Fig. 1b). This SHIP-only set of seasonal sampling experiments is our baseline as it is also used in all platform combinations. Three seasonal sampling scenarios (summer (smr), summer+winter (smr+wtr), and autumn+spring (aut+spr)) were considered. While the first two scenarios are addressed in detail in this study (Fig. 5a-b, and Table 1), the third one can be found in the supplementary material (Fig. S3, and Table S4) in support of the main points already made in Fig. 5a-b.

https://doi.org/10.5194/bg-2021-344
Preprint. Discussion started: 3 January 2022

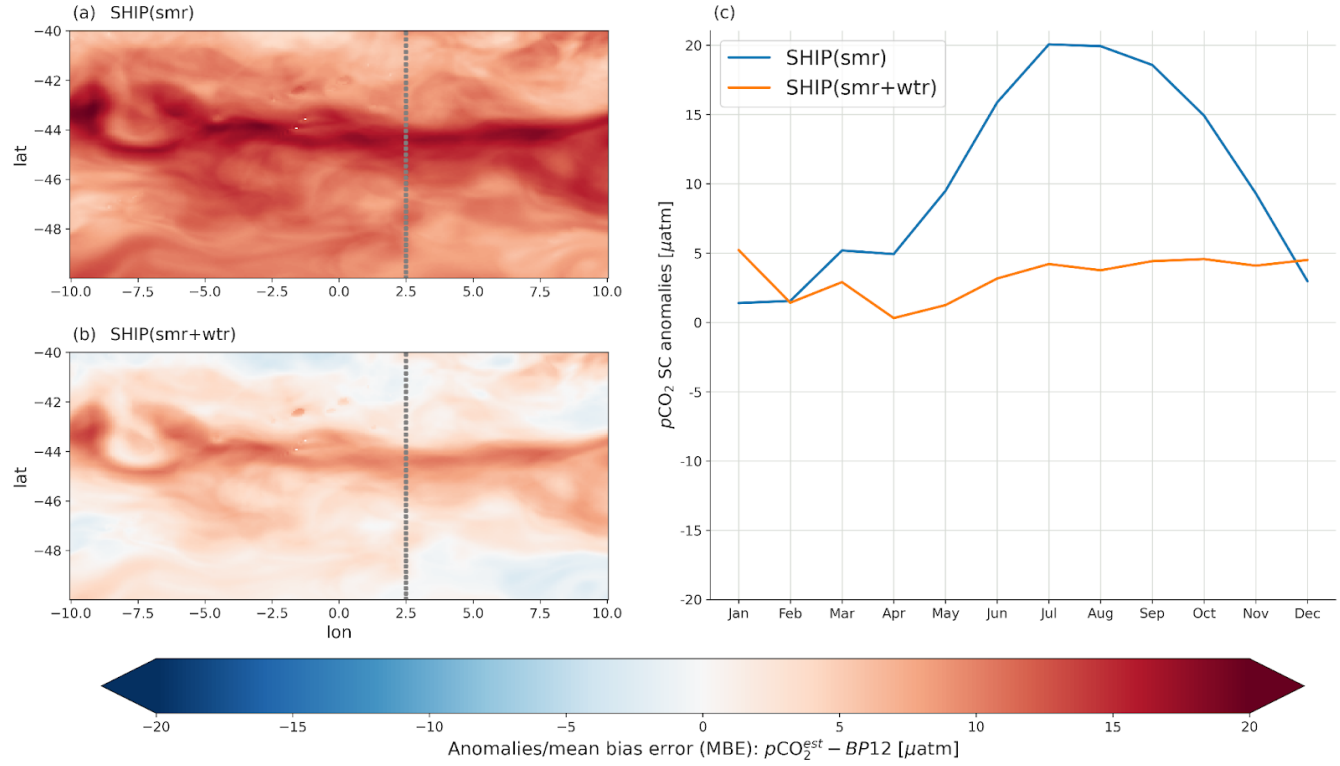


Figure 5: Reconstruction anomalies for the idealized SHIP experiment where the idealized ship sampled the domain based on the sampling regimes/scenarios, SHIP(smr) for summer, SHIP(smr+wtr) for summer and winter. Panels (a) and (b) show the maps of the reconstruction anomalies according to the two sampling regimes SHIP(smr), and SHIP(smr+wtr) respectively; panel (c) shows the anomalies of the mean seasonal cycle (SC) reconstruction based on these two sampling regimes; that is, SHIP(smr) and SHIP(smr+wtr). The meridional dotted grey line in panels (a) and (b) illustrates the sampling line (summer & winter) and serves as a reminder of how SHIP sampling was performed.

The spatial and seasonal cycle anomalies from the reconstructions for the summer (smr), summer and winter (smr+wtr) sampling lines are depicted in Fig. 5a-b. The results for the autumn and spring (aut+spr) sampling lines are summarized in the supplementary materials (Fig. S3). The uncertainties and regression errors for all three experiments are shown in Table 1. These results showed that the highest positive anomalies in the reconstruction of the mean and the seasonal cycle occurs when a ship samples (i.e. makes 2 passes in consecutive months) the sub-domain only in summer (Fig. 5a, c). This sampling strategy resulted in a strong positive anomaly (±20 µatm) that peaks in winter and weakens in mid-summer (Fig. 5c). In sharp contrast, when winter sampling crossings are added to the summer scenario (smr+wtr) the spatial and seasonal anomalies are significantly reduced from 20 µatm to < 5 µatm respectively (Fig. 5b, c). The weaker but persistent positive anomaly in the SAF accounts for most of the reduced positive seasonal cycle anomaly (Fig. 5a, c).

All scenarios depict a mesoscale modulated positive annual pCO_2 anomaly (MBE) climatology in the vicinity of the SAF (Fig. 5a-b). However, whereas this is slightly offset by equally strong positive anomalies in the SAZ and PFZ for the smr scenario

https://doi.org/10.5194/bg-2021-344
Preprint. Discussion started: 3 January 2022

(Fig. 5a), the meridional gradients of the anomalies are much weaker for the smr+wtr scenario (Fig. 5b). These differences are very well reflected in the anomalies of their corresponding seasonal cycles (Fig. 5c).

Table 1: ML regression modelling scores of the ensemble average (ML2) for two sampling scenarios of SHIP experiment: SHIP(smr) for summer sampling, and SHIP(smr+wtr) for summer and winter sampling. The configuration of this set of experiments is presented in Table S2 and clearly described in Sect. 2.3.4. The first column of the table is the experimental set and the second one corresponds to the considered experiments. The statistical metrics used to assess ML2 for this set of experiments are abbreviated as follows: RMSE is the root mean square error calculated following Eq. (4); MAE is the mean absolute error (Eq. 3); MBE or Bias is the mean average error (Eq. 2); and r is the Pearson's correlation coefficient (Eq. 5) between the reconstructed and BP12 model truth pCO_2. Values in the table are significantly different from the mean for the corresponding column (with 95% confidence level or p-value < 0.05 for the two-tailed Z-test).

Set	Experiments	RMSE (µatm)	MAE (µatm)	MBE (µatm)	r
SHIP	SHIP(smr)	13.79	11.51	10.52	0.36
	SHIP(smr+wtr)	6.8	5.29	3.18	0.73

These SHIP-only experiment results (Tables 1 and S3) also show that the summer-only sampling of the sub-domain both produces the largest sampling bias (10.52 µatm, with an RMSE of 13.79 µatm) and yields the weakest correlation between the underlying pCO_2 estimates and the model ground-truth (with r = 0.36). On the other hand, it also showed that if the ship undertakes just one more meridional voyage in winter, this halved the RMSE to 6.8 µatm and the bias (MBE) to 3.18 µatm compared to the summer-only sampling experiment, SHIP(smr). Moreover, it also strengthened the linear association between the reconstruction and BP12 model ground-truth for pCO_2 (r = 0.73).

3.2.2 Idealized SHIP and autonomous observations platform experiments

In this section, we presented the results of three sets of combined ship and autonomous platform experiments (SHIP(smr) + FLOAT, SHIP(smr) + WG, and SHIP(smr) + nUSV) that allowed us to test the hypothesis that complementing summer biased ship-based sampling with year-long high-resolution sampling in space and time reduces the reconstruction uncertainties and positive annual mean and seasonal cycle biases relative to the ship-sampling alone (Figs. 5a, c, and 6a-b)(Gregor et al., 2019; Bushinsky et al., 2019; Sutton et al., 2021). We simulated and analysed the reconstruction of the mean annual pCO_2 and seasonal cycles from carbon-floats (FLOAT) and carbon Wavegliders (WG) deployed independently for a year in the Sub-Antarctic Zone (SAZ) and Polar Frontal Zone (PFZ) (Figs. 5b, 5c-d, and 5e-f). These were complemented by simultaneous year-round FLOAT deployments in the SAZ and PFZ (Fig. 5b, g), and a deployment of the new unmanned surface vehicle (nUSV) Saildrone that spanned across all three domains (Fig. 5b, h).

https://doi.org/10.5194/bg-2021-344
Preprint. Discussion started: 3 January 2022

These results show that both the reconstructed mean annual anomaly and the seasonal cycle of pCO_2 are very sensitive to the spatial and temporal characteristics of the additional autonomous sampling platform (Fig. 6). Statistics (Table 2) show that all the autonomous platform deployments experiments improved the significant winter positive biased seasonal cycle anomaly from the summer ship sampling reconstruction (±20 µatm). However there remained a small but variable (2 - 10 µatm) winter - spring seasonal bias in all deployment combinations (Fig. 6b). The exception was the experiment with a FLOAT deployment in the SAZ, which resulted in a negative seasonal bias that also peaked in winter (±10 µatm) and started earlier in the autumn (Fig. 6b). The two experiments with the smallest seasonal biases were the SHIP(smr) + WG (SAZ), and SHIP(smr) + nUSV. The first, SHIP(smr) + WG(SAZ), showed a small negative bias in the summer (< -5 µatm) and a small positive bias in the winter (< 5 µatm). The latter, SHIP(smr) + nUSV, showed a small positive bias in summer (0 - 5 µatm) and in winter (4 - 5atm)(Fig. 6b). In contrast, the experiment SHIP(smr) + FLOAT(SAZ+PFZ) that combined the two year-round FLOAT deployments (SAZ and PFZ), shows a minimal bias in summer but among the highest for all the experiments in winter (±10 µatm)(Fig. 6b).

The spatial annual mean pCO_2 experimental scenario anomalies are consistent with the characteristics of the seasonal cycle of pCO_2 (Fig. 6a, 6c-h). In all cases the Sub-Antarctic Front (SAF) emerged as a feature with a variable positive pCO_2 anomaly relative to the SAZ and PFZ sectors to the north and south respectively (Fig. 6a, 6c-h). All the scenarios highlight significant mesoscale anomaly gradients across all the domains (Fig. 6a, 6c-h). The year-long deployment of FLOATs and WGs in the SAZ lead to negative anomalies in both the SAZ and the PFZ but those for the WG experiments are significantly weaker (Fig. 6c, e).

https://doi.org/10.5194/bg-2021-344
Preprint. Discussion started: 3 January 2022

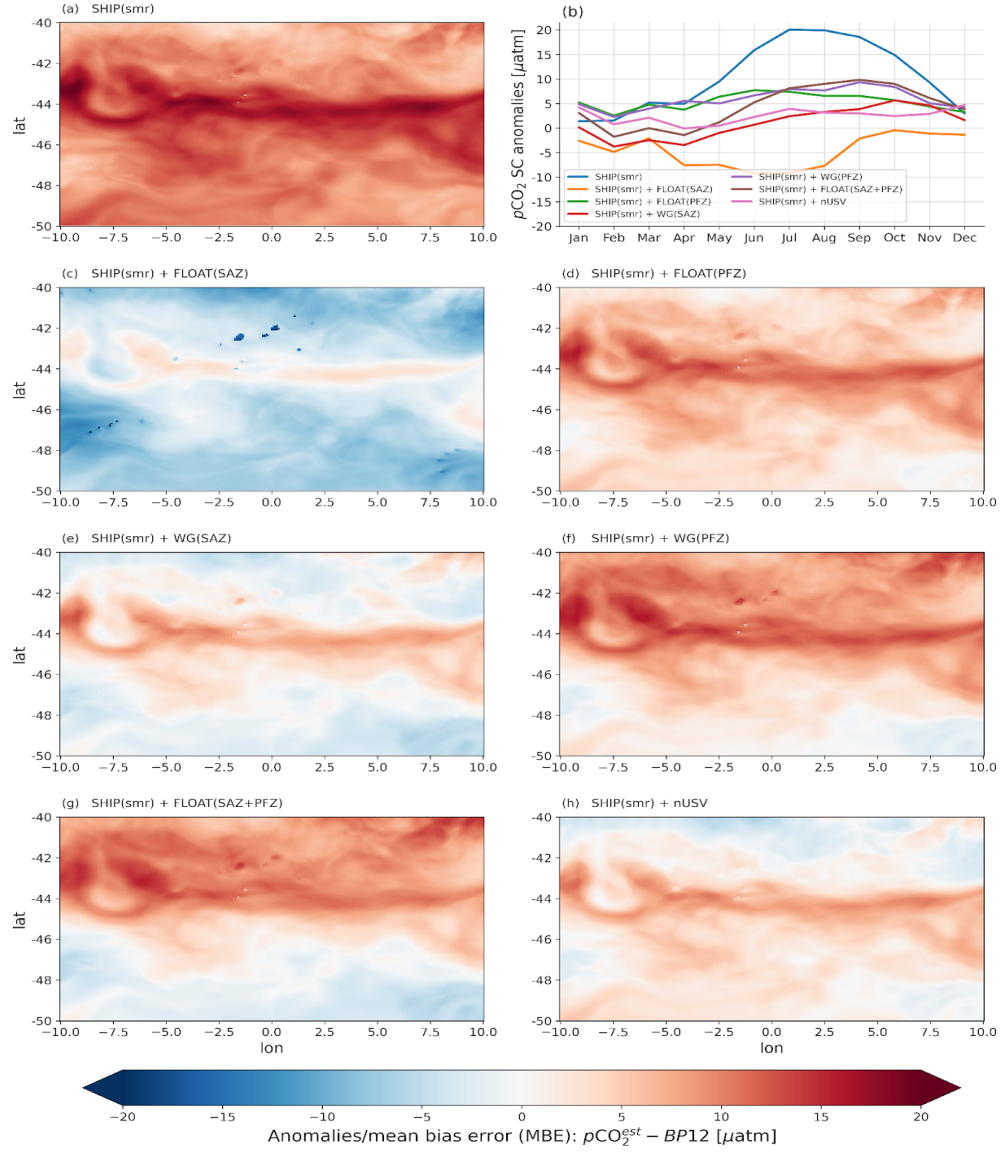


Figure 6: Reconstruction anomalies for the 4 sets of experiments, SHIP, SHIP + FLOAT, SHIP + WG, and SHIP + nUSV with a particular focus on the summer-only baseline scenario: SHIP(smr). Panels (a) shows the spatial anomalies or biases (MBEs) of the mean annual pCO_2 reconstruction for the SHIP summer-only sampling scenario, that is, SHIP(smr); panel (b) shows the anomalies of mean seasonal cycle (SC) reconstructions of the summer-only sampling scenario of the above-mentioned sets of experiments; panels (c) and (d) show the spatial reconstruction anomalies for the SHIP + FLOAT experiments where two independent FLOATs were deployed in the SAZ and PFZ (respectively), and used to supplement SHIP(smr); panel (e) and (f) show the spatial reconstruction anomalies for the SHIP + WG experiments where two independent WGs were deployed along the SHIP line in the SAZ and PFZ (respectively), and used to supplement SHIP(smr); panel (g) shows the spatial anomalies of the mean annual pCO_2 reconstruction for the SHIP + FLOAT experiment scenario where the two FLOAT deployments (SAZ and PFZ) were used to supplement the SHIP summer-only sampling scenario, hence SHIP(smr) + FLOAT(SAZ+PFZ); and panel (h) shows the spatial anomalies of the mean annual pCO_2 reconstruction for the SHIP + nUSV experiments where SHIP(smr) were supplemented by a year-round sampling of the nUSV Saildrone, hence SHIP(smr) + nUSV.

https://doi.org/10.5194/bg-2021-344
Preprint. Discussion started: 3 January 2022

However, the reverse was found for the SAF zone which shows a stronger positive anomaly for the WG(SAZ) than for the FLOAT(SAZ) (Fig. 6c, e). The stronger mean annual negative pCO_2 anomaly for the SHIP(smr) + FLOAT(SAZ) deployment is consistent with the negative seasonal cycle anomaly, which points to the mean annual anomaly being mainly influenced by the winter negative anomaly (Fig. 6b-c). Similarly, the much weaker negative anomalies in the SAZ and PFZ for the WG deployment are consistent with the weaker seasonal cycle (< ±5 µatm) of pCO_2 for the whole domain.

SHIP(smr) + FLOAT(PFZ) and SHIP(smr) + WG(PFZ) deployments result in weak to moderate positive anomalies in the northern half of the PFZ, the SAZ and the SAF and weak to zero anomalies in the southern PFZ, all of which are characterized by mesoscale gradients (Fig. 6d, f). Both scenarios show a comparable positive seasonal cycle anomaly although the phasing of the winter maximum is earlier Jun vs Sep for the SHIP(smr) + FLOAT(PFZ) (Fig. 6b). The mean annual pCO_2, from the combined SHIP(smr) + FLOAT(SAZ+PFZ) deployments, showed spatial characteristics similar to the SHIP(smr) + FLOAT(PFZ) but with intensified negative and positive anomalies in the PFZ and SAZ respectively (Fig. 6g). The moderately strong positive winter anomalies (±10 µatm) in the seasonal cycle for this experiment indicate that the mean annual positive anomalies are also dominated by the winter anomalies (Fig. 6b). The mean annual pCO_2 anomaly for the SHIP(smr) + nUSV deployments is weakly negative (< -5 µatm) in the north SAZ and weakly positive (< 5 µatm) in the SAF and the PFZ (Fig. 6h). The overall weak mean annual pCO_2 anomaly is consistent with the weakest (0 – 5 µatm) seasonal cycle anomaly (Fig. 6b).

Table 2: ML regression modelling scores of the ensemble average (ML2) for the summer-only sampling scenario (smr) of all the 4 sets of experiments: SHIP, SHIP + WG, SHIP + FLOAT, and SHIP + nUSV. The configuration of these experiments is presented in Table S2 and described in Sect. 2.3.4. Similar to Table 1, the first column of the table is the experimental set and the second one corresponds to the considered experiments. The statistical metrics used to assess ML2 for this set of experiments are abbreviated as follows: RMSE is the root mean square error calculated following Eq. (4); MAE is the mean absolute error (Eq. 3); MBE or Bias is the mean average error (Eq. 2); and r is the Pearson's correlation coefficient (Eq. 5) between the reconstructed and the BP12 model truth pCO_2. Values in the table are significantly different from the mean for the corresponding column (with 95% confidence level or p-value < 0.05 for the two-tailed Z-test).

Sets	Experiments	RMSE (µatm)	MAE (µatm)	MBE (µatm)	r
SHIP	SHIP(smr)	13.79	11.51	10.52	0.36
SHIP + FLOAT	SHIP(smr) + FLOAT(SAZ)	9.29	7.46	-4.81	0.60
	SHIP(smr) + FLOAT(PFZ)	8	6.51	5.32	0.73
	SHIP(smr) + FLOAT(SAZ+PFZ)	9.12	7.57	4.14	0.63
SHIP + WG	SHIP(smr) WG(SAZ)	6.88	5.4	0.82	0.64
	SHIP(smr) WG(PFZ)	9.41	7.59	5.88	0.57
SHIP + nUSV	SHIP(smr) + nUSV	6.4	5.1	2.38	0.74

https://doi.org/10.5194/bg-2021-344
Preprint. Discussion started: 3 January 2022

Table 2 shows that SHIP(smr), the baseline biased ship-summer sampling experiment (the status quo in the Southern Ocean) yielded an RMSE of 13.79 µatm and a mean biased error of 10.52 µatm which is comparable with the Southern Ocean results for CSIR-ML6 (Gregor et al., 2019). Table 2 also shows that although all the additional high resolution platform experiments reduced the RMSE and MBE, the magnitude of the impact was very sensitive to the platform and its location. All three scenarios of the year-long SHIP(smr) + FLOAT experiments reduced the RMSE of SHIP(smr) experiment by 32.6 - 41.9% however, only the scenario SHIP(smr) + FLOAT (PFZ) provided the lowest RMSE and MAE as well as the statistically significant correlation (r = 0.73) between the estimates and known truth. Both WG experiments (SAZ and PFZ deployments) also reduced the RMSE by 31.7 - 50.1% through a statistically significant correlation with r = 0.64 (SAZ) and r = 0.57 (PFZ), respectively (Table 2). The SHIP(smr) + nUSV experiment yielded the lowest RMSE (6.4 µatm)(53.5%), MAE and MBE with a significant correlation with r = 0.74. These results are consistent with the comparative seasonal cycle anomalies that showed the SHIP(smr) + FLOAT(PFZ) and SHIP(smr) + nUSV to have the smallest seasonal cycle biases (Fig. 6b), and higher correlations with the known truth (with r = 0.73 and r = 0.74 respectively).

4 Discussion

Resolving the variability and trends of the seasonal cycle of pCO_2 in the Southern Ocean has been a long-term objective for the ocean carbon community in order to reduce the uncertainties and biases of the seasonal and mean annual fluxes (Lenton et al., 2006, 2013; Takahashi et al., 2009; Monteiro et al., 2015; Mongwe et al., 2018; Gregor et al., 2018; Bushinsky et al., 2019; Sutton et al., 2021). This started with largely observation-based approaches which constrained the seasonal cycle climatology (Takahashi et al., 2009, 2012) and set requirements to resolve the variability (Lenton et al., 2006; Monteiro et al., 2015). The advent of a globally coordinated surface ocean CO_2 data, SOCAT (Bakker et al., 2016), together with machine learning methods (Landschützer et al., 2014, 2016; Rödenbeck et al., 2015) provided a basis for spatial and temporal gap filling that has resulted in an internally consistent set of reconstructions for the ocean and Southern Ocean CO_2 fluxes that contribute to the global carbon budget (Canadell et al., 2021; Fay et al., 2021; Friedlingstein et al., 2021).

However, Gregor et al. (2019) argued that the uncertainties and biases of CO_2 flux reconstructions are now limited by both data gaps and variability-scale sensitivity of surface ocean CO_2 observations – a boundary that the authors dubbed “the wall”. Our results make the key point that the seasonal and mean annual biases and uncertainties (RMSEs) in the reconstructions depend critically on simultaneously resolving the spatial, meridional gradients, and temporal, seasonal and intra-seasonal variability. We now discuss three sampling scale sensitivities emerging from our analysis and what we suggest is required to get "over the wall": (1) the sensitivity of the reconstructions to the seasonal cycle, (2) the sensitivity of the reconstructions to the seasonal cycle of the meridional gradients, (3) the sensitivity of the reconstructions to the intra-seasonal variability, (4) the

https://doi.org/10.5194/bg-2021-344
Preprint. Discussion started: 3 January 2022

need to simultaneously sample the meridional gradients and their intra-seasonal variability to get "over the wall", and (5) the limitations of this study.

4.1 Seasonal sampling scale sensitivity

The SHIP-only sampling experiments, which most closely simulate the historical ship-based and seasonally biased SOCAT gridded database in the Southern Ocean, point towards an unexpectedly high sensitivity of the reconstruction uncertainties and biases to the seasonal sampling scales (Figs. 1b and 5a-b, and Table 1). Simulation of the existing Southern Ocean ship-summer sampling, SHIP(smr), resulted in a seasonal cycle reconstruction with a strong positive winter out-gassing seasonal anomaly bias of ±20 µatm that was strong enough to reverse the in-gassing flux from the model domain (Fig. 5c), which also biased (positively) the spatial mean annual flux for the domain (Fig. 5a). The impact of the biased summer sampling is also expressed in the comparatively elevated RMSE: 13.79 µatm (Table 1), which is of a magnitude close to the RMSEs of the ML methods for the Southern Ocean - particularly in the Polar Frontal Zone (PFZ). For example, (Gregor et al., 2018) reported in the PFZ an average RMSE value of 14.33 µatm, and also RMSE = 13.09 µatm for the SOM-FNN method (Peter Landschützer et al., 2016) within the same region (PFZ). Furthermore, a comparative analysis of the SHIP summer-only experiment, SHIP(smr), and the SHIP summer and winter one, SHIP(smr+wtr), shows that SHIP(smr+wtr) outperformed SHIP(smr) across all the performance metrics (Table 1) by halving them, for instance, RMSE = 6.88 µatm. The sensitivity of the reconstruction to the seasonal sampling bias is again further emphasised by the impact of the addition of a single SHIP meridional 2-leg winter (Jul-Aug) sampling lines, SHIP(smr+wtr), which reduced the mean monthly winter anomaly of pCO_2 for the whole domain from ±20 µatm in winter to less than 5 µatm over the whole seasonal cycle (Fig. 5a-c). The impact of the additional winter line is also expressed in the reduction of the bias error from (Table 1).

When splitting the anomalies across the two sub-domains (SAZ and PFZ) for the SHIP(smr) scenario, a comparable seasonal sampling bias sensitivity was found for the SAZ and PFZ domains (Fig. 7a). The winter reconstruction bias dominates any internal variability in the sub-domains. However, the introduction of the SHIP-winter line not only impacted on the overall mean seasonal bias but also shows that the mean seasonal cycle comprises out-of-phase seasonal modes of variability in both SAZ and PFZ domains (Fig. 7b).

https://doi.org/10.5194/bg-2021-344
Preprint. Discussion started: 3 January 2022

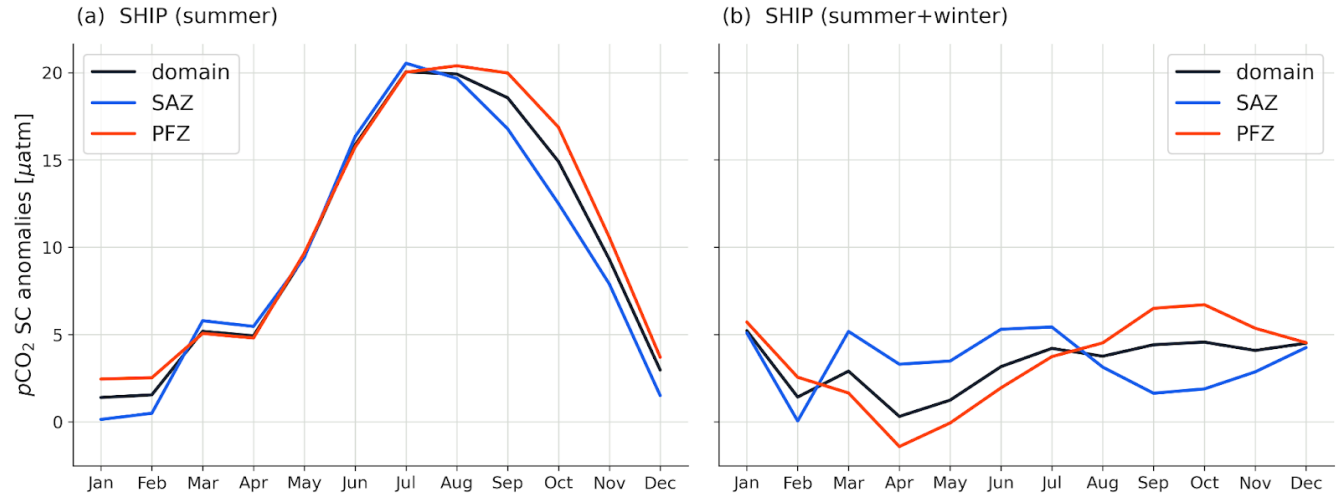


Figure 7: Anomalies of the mean surface ocean pCO_2 seasonal cycle (SC) reconstructions from two SHIP-only experiments. Panel (a) shows the pCO_2 SC anomalies from the SHIP (summer-only)-based reconstruction in the whole domain, the SAZ, and the PFZ; and in contrast, panel (b) shows the pCO_2 SC anomalies from the SHIP (summer + winter)-based reconstruction for the whole domain, the SAZ, and the PFZ.

It suggests that an important outcome of the reduction of seasonal and mean biases is the emergence of important modes of variability that can provide a useful window into key processes as well as identifying key modes of variability that can influence sampling strategies (Fig. 7a-b). Our findings on the sampling bias sensitivity are consistent with the early estimates of the minimum number of ship transects required to observationally resolve the seasonal cycle in the Southern Ocean as being quarterly, across the 4 seasons, and zonally 30º apart (Lenton et al., 2006; Monteiro et al., 2010). Together with these early results our analysis confirms that additional ship pCO_2 observation lines in summer will not be a useful contribution towards reducing the uncertainties and biases of the reconstructions. Rather, as proposed earlier, additional seasonal sampling lines in winter will make a decisive impact (Figs. 5a-c and 7a-b, and Table 1). However, realistically this is not achievable because access to the Southern Ocean outside the summer period is logistically challenging outside the Drake Passage (Monteiro et al., 2015; Gray et al., 2018).

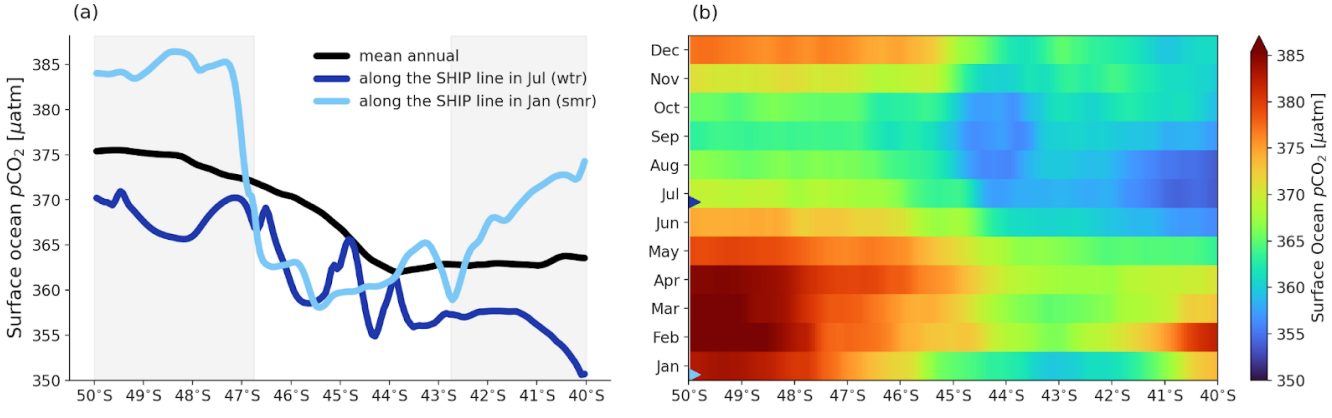
The well recognized seasonal sampling bias problem, outside the Drake Passage (Munro et al., 2015), is being addressed globally and in the Southern Ocean using a variety of autonomous sampling platforms such as Wavegliders, pH-Floats, and Saildrones (Monteiro et al., 2015; Williams et al., 2017; Gray et al., 2018; Bushinsky et al., 2019; Sutton et al., 2021). We now discuss the effectiveness of each one through experiments to simulate their sampling characteristics inside the model domain. All these experiments include the SOCAT-like SHIP-summer observations. These experiments focus primarily on the impact of the autonomous sampling platforms WG and pH-Floats as both have been deployed in the Southern Ocean with sampling strategies that view to address the seasonal sampling bias (Monteiro et al., 2015; Gray et al., 2018; Gregor et al., 2019). We return to the potential of Saildrones later in the discussion in the context of a discussion of how to "get over the wall".

https://doi.org/10.5194/bg-2021-344
Preprint. Discussion started: 3 January 2022

Biogeosciences Discussions Open Access EGU

4.2 The seasonal cycle of the meridional gradients

One of the unexpected results from our analysis was that the ship-based reconstruction with both summer and winter crossings of the domain, SHIP(smr+wtr), performed as well as the best reconstructions in which the SHIP summer-only sampling, SHIP(smr), is supplemented with an autonomous vehicle WG or FLOAT sampling continuously throughout the year (Tables 1 and 2). Thus, SHIP(smr+wtr) performed better (e.g., RMSE = 6.8 µatm) than the SHIP(smr) + FLOAT(PFZ) and SHIP(smr) + WG(SAZ) experiments that produced RMSEs of 8.0 µatm and 6.88 µatm, respectively. These results suggest that while resolving the local seasonal cycle of the surface ocean pCO_2 with the WG and the FLOATs had a decisive impact on the RMSEs and mean biases (MBEs), an additional scale is being resolved by the SHIP experiment in winter, which is not addressed by the sampling scales of the two autonomous sampling platforms WG (1-day period) and FLOAT (10-day period). Here, we propose that the critical missing scale is the variability of the meridional gradient of surface ocean pCO_2 (Fig. 8a), or more critically, the seasonal cycle of the meridional gradient of pCO_2 (Fig. 8b). Together these figures highlight that although the mean increasing southward gradient in pCO_2 is sustained throughout the annual cycle (Fig. 8a), there are sharp seasonal spatial and temporal contrasts in the meridional variability of the magnitudes (Fig. 8b). This includes significant seasonal differences in the influence of mesoscale on the spatial variability (Fig. 8a). The climatological meridional gradients of surface DIC and pCO_2 in the Southern Ocean are well characterized through *in situ* observations (Wu et al., 2019), data-products (Gregor et al., 2018, 2019) and models (Hauck et al., 2015, 2020). These results highlight that characterizing the meridional gradient is not sufficient in itself because shipboard observations in the SOCAT database already include the meridional gradients but these observations in the Southern Ocean are strongly biased toward summer (Gregor et al., 2019; Gregor and Gruber, 2021). As our study indicates, the seasonal scale variability of that meridional gradient matters the most, which is why SHIP(smr+wtr) makes such a difference (Tables 1 and 2) compared to SHIP(smr) + WG and SHIP(smr) + FLOAT.



https://doi.org/10.5194/bg-2021-344
Preprint. Discussion started: 3 January 2022

Figure 8: Seasonal contrasts for the meridional gradient (MG) of surface ocean pCO_2 in the experimental sub-domain. Panel (a), shows the mean annual MG (black), the mean MG along the SHIP line in summer (January)(light blue) and in winter (July)(dark blue); and panel (b) shows the seasonal cycle of the meridional gradient of pCO_2 showing the months when the SHIP sampled (blue triangle markers) with the light blue for Jan (smr) and the dark blue for Jul (wtr). The light grey shadings in panel (a) show the sub-domain areas (north and south) where large differences in pCO_2 meridional gradients along the SHIP line in summer and winter.

Significant differences exist between the meridional gradients along the SHIP line in summer (e.g., January) and winter (e.g., July) (Fig. 8a-b). For example, these differences are more significant farthest south (> 47ºS) and farthest north (< 43ºS) compared to the middle (43ºS - 47ºS) of the sub-domain (light grey shadings, Figs. 8a). Similarly, the seasonal cycle difference is not as big in the middle of the sub-domain as it is at the extreme lines of the SAZ and PFZ (Fig. 8b). That is why we need a sampling platform that is able to capture critical scales of variability. Another key point we raised concerning the sampling scale sensitivity on the pCO_2 reconstructions is that resulting uncertainties and biases depend on the seasonal scale of the meridional gradients of the surface ocean pCO_2 (Fig. 8b). Shedding light of this point results in resolving the seasonal cycle of the meridional gradients.

The similarity of the anomalies between the SHIP(smr) + WG(SAZ) and SHIP(smr) + nUSV experiments are supported by the impact that these sampling strategies have on the seasonal cycle of the bias (Fig. 6b). This shows that, relative to other sampling experiments, there was a reduction of the biases across the whole seasonal cycle but more so in summer-autumn and less so in winter-spring (Fig. 6b). The significantly smaller MBE for SHIP(smr) + WG(SAZ) can be ascribed to the bias being slightly negative in summer-autumn and positive in winter-spring which leads to a small mean annual MBE whereas in the case of the SHIP(smr) + WG(PFZ) experiment, the MBE is small but positive throughout (Fig. 6b, and Table 2). The mean annual anomaly map of pCO_2 for the SHIP(smr) + nUSV experiment still shows a positive anomaly, though weaker, at the frontal zone because although the nUSV Saildrone has a daily sampling resolution, it is only crossing the highly synoptic SAF zone periodically (Fig. 1c). This is consistent with all the instances when not resolving the temporal variability results in a positive bias of varying magnitudes (Fig. 6b).

On designing an observation-based strategy for quantifying the Southern Ocean uptake of CO_2, Lenton et al. (2006) argued that constraining the net seasonal air-sea CO_2 fluxes within the natural variability of the carbonate system requires doubling the current Southern Ocean meridional sampling. In a semi-idealized experimental setting, our study takes this further by showing that resolving the seasonal cycle of the meridional gradients is very critical. WG and FLOAT provide high temporal sampling resolution but they do not resolve the existing meridional gradients. Therefore, increasing data density through zonal autonomous sampling vehicles (e.g., floats) is not sufficient to minimize reconstruction errors. The quarterly meridional sampling strategies proposed by Lenton et al. (2006) and Monteiro et al. (2010) could help to resolve the seasonal cycle of the meridional gradients, but they are not operationally feasible.

https://doi.org/10.5194/bg-2021-344
Preprint. Discussion started: 3 January 2022

4.3 Intra-seasonal variability of the Seasonal Cycle

Recent high-resolution observations using different types of carbon-enabled autonomous platforms have highlighted a potential sensitivity of Southern Ocean CO_2 flux reconstruction uncertainties and mean bias to aliases in sampling the intra-seasonal to seasonal temporal scales (Monteiro et al., 2015; Williams et al., 2017; Gray et al., 2018; Bushinsky et al., 2019; Sutton et al., 2021). Here we discuss the sensitivity of the model domain reconstruction statistical metrics to a range of semi-idealized scenarios of SHIP-summer supplemented with FLOAT and WG observations (Table 2; Fig. 6). The nUSV, Saildrone analogue, sampling scenario is brought in later to test the predicted sampling requirements to achieve the lowest RMSEs and mean bias error. There was no real benefit in reproducing the zonal sampling approach for the Saildrone (Sutton et al., 2021) because it would be comparable to the zonal travel of FLOAT but with higher daily sampling more akin to the WG. Its metrics would therefore have been comparable to both and contributed little to learning.

One of the standout aspects of this part of the analysis, investigating the impact of sampling period, was the significant difference in the uncertainty and biases between the best performing SHIP(smr) + WG(SAZ) (RMSE = 6.88 µatm; MBE = 0.82 µatm) and SHIP(smr) + FLOAT(PFZ) (RMSE = 8µatm; MBE = 5.32 µatm) scenarios (Table 2). These comparative statistics point to the reconstructions also being very sensitive particularly to the temporal sampling scales. This finding can be explained and understood from the characteristics of the variability from time series from single model grid cells in the SAZ, on the SAF, and in the PFZ (Fig. 9). Local scale single grid-cell observations are appropriate instead of spatial means because they simulate the local nature of the variability and how it is observed. The variability characteristics of these time series help explain the statistics of the pCO_2 reconstructions (Fig. 9; Table 2). The SAZ and SAF are characterized by stronger intra-seasonal variability whereas the PFZ is characterized by lower frequency (sub-seasonal) - seasonal modes of variability (Fig. 9). Thus, while the SAZ and SAF sub-domains and their stronger intra-seasonal variability are best resolved by the daily sampling of the WG, the PFZ domain, which is dominated by the lower frequency sub-seasonal to seasonal cycle, is resolved equally well by the WG - daily and FLOAT - 10-daily sampling periods (Fig. 9; Table 2).

https://doi.org/10.5194/bg-2021-344
Preprint. Discussion started: 3 January 2022

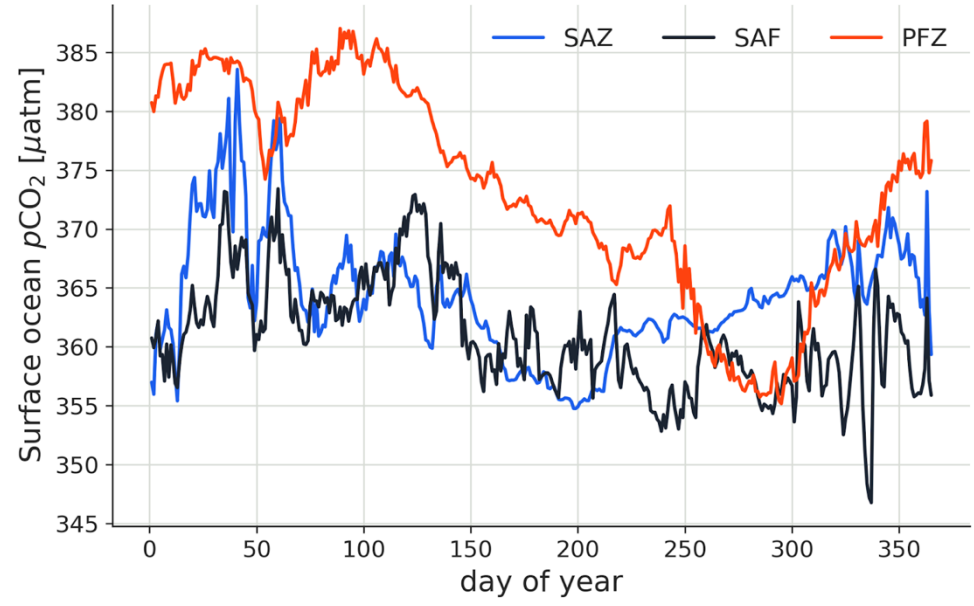


Figure 9: Time series (one year) plots of the variability of surface ocean pCO_2 at single model grid cells on the SHIP line (2.5ºE, Fig. 1b). We used the following single model grid cells: 42ºS, 2.5ºE in the Sub-Antarctic Zone (SAZ); 44ºS, 2.5ºE on the Sub-Antarctic Front (SAF); and 47ºS, 2.5ºE in the Polar Frontal Zone (PFZ). It shows that while the SAZ and SAF are dominated by synoptic modes of variability, the PFZ is characterized by longer period sub-seasonal to seasonal scales of variability.

Therefore, given that WGs and FLOATs in these sampling scenarios are comparable in that neither have a strong meridional gradient resolving sampling strategy, the main difference between them is the daily sampling rate of the WGs and the 10-day sampling rate for the FLOATs. Fig. 9 then helps explain why even though the domain reconstructions based on FLOAT(PFZ) sampling scenario perform best out of the two FLOAT scenarios, SHIP(smr) + FLOAT(SAZ+PFZ), ultimately it underperformed relative to the WGs because it was aliasing the synoptic intra-seasonal variability in the SAZ and SAF. The finding that the high temporal resolution of the SHIP(smr) + WG(SAZ) was the only SHIP(smr) + WG and SHIP(smr) + FLOAT combination to match the performance of the SHIP(smr+wtr) experiment whose strength was in resolving the seasonal contrasts of the spatial meridional gradient, suggests that these two scales of variability are close to equally important towards a low bias and RMSE reconstruction. Resolving the former and the latter simultaneously may therefore be a critical step.

More broadly and relative to the SHIP summer-only scenario, all the annual cycle experiments yielded a reduction in the reconstructed seasonal cycle anomalies (Fig. 6b), in the uncertainties (32 - 50%), biases (±50%) as well as a statistically significant improvement for Pearson's correlation coefficient (r) (Fig. 6a-b, and Table 2). When comparing SHIP(smr) + WG with SHIP(smr) + FLOAT, reconstructed annual mean pCO_2 maps for the whole domain were consistent with reduced anomalies, for instance, with small positive anomalies for SHIP(smr) + FLOAT(PFZ) and small negative anomalies for SHIP(smr) + WG(SAZ) (Fig. 6d-e, respectively). However, while comparing SHIP(smr) + WG(SAZ) with SHIP(smr) +

https://doi.org/10.5194/bg-2021-344
Preprint. Discussion started: 3 January 2022

FLOAT(SAZ) where the WG and FLOAT are both deployed in the SAZ, there is a significant difference in the RMSEs and MBEs with respectively 6.88 µatm and 0.82 µatm for the former, and 9.29 µatm and -4.81 µatm for the latter (Table 2).

This analysis provides additional understanding into the strengths and limitations of the way that the 3 main autonomous platforms (Wavegliders, carbon-floats and Saildrones) deployed in the Southern Ocean contribute to increasing or decreasing the seasonal cycle and mean annual biases as well as the RMSEs (Monteiro et al., 2015; Bushinsky et al., 2019; Sutton et al., 2021). On the basis of daily mean resolution of the pCO_2, Monteiro et al. (2015) showed that a temporal sampling resolution of less than 2 days would be necessary in 30 - 40% of the Southern Ocean, corresponding to the SAZ, in order to reduce the uncertainty to less than 10% of the annual mean. Our study confirms the sensitivity of the RMSE of the intra-seasonal variability sampling alias and also shows that it impacts on the bias of the annual mean. SOCCOM-float calculated pCO_2 data has made a decisive impact on resolving the seasonal cycle in the Southern Ocean and suggests that winter CO_2 out-gassing may be underestimated in SOCAT based reconstructions (Gray et al., 2018; Bushinsky et al., 2019). Our study suggests that this observed and reconstructed elevated out-gassing fluxes may be the result of both aliasing of the intra-seasonal variability as well as not resolving the seasonal cycle of the meridional gradient. Our analysis also raises a question around the assumption that not resolving the intra-seasonal variability of pCO_2 does not contribute significantly to the RMSE and the bias (Bushinsky et al., 2019). It shows that the intra-seasonal modes of the wind are not sufficient to impart a low mean annual and seasonal cycle bias. We propose that the impact of SOCCOM floats on the reconstructions can be improved by reducing the sampling period to 1 - 2 days and a coordinated meridional deployment that helps to resolve the meridional gradient across the annual cycle. Our study also suggests that notwithstanding the high temporal frequency of the USV Saildrone, the present emphasis on a zonal sampling pattern (Sutton et al., 2021) also underestimates the potential contribution that this platform could make in observing the seasonal cycle of the meridional gradient at high temporal resolution simultaneously. We now examine this aspect in more detail.

4.4 “Getting over the wall” in the Southern Ocean by resolving the intra-seasonal and seasonal variability of the meridional gradient - Proposed optimal sampling strategy

This analysis has highlighted that in order to minimize the uncertainties and biases sufficiently to "get over the wall”, observational strategies in the Southern Ocean need to simultaneously resolve the seasonal cycle of the meridional gradient at temporal scales that also resolve, where necessary, the intra-seasonal variability. To test this hypothesis, we designed an additional year-round deployment experiment that simulated the sampling capabilities of the new unmanned surface vehicle (nUSV) Saildrone (Sutton et al., 2021) to supplement the SHIP summer-only sampling SHIP(smr) (Figs. 1c and 6b, h); that is, SHIP(smr) + nUSV. This experiment combined the speed of the nUSV Saildrone (Meinig et al., 2019; Gentemann et al., 2020) required to cover the regional meridional spatial gradients length scales (Fig. 1c), with high frequency daily sampling to supplement SHIP(smr). Together these fulfill the requirements that emerged from the earlier analysis.

https://doi.org/10.5194/bg-2021-344
Preprint. Discussion started: 3 January 2022

730

Comparative statistics shows that SHIP(smr) + nUSV experiment yielded a very significant improvement of the reconstruction skills relative to all other platform combinations (Table 2). Its performance metrics (RMSE = 6.4 µatm) outperformed the next best combination SHIP(smr) + WG(SAZ) (RMSE = 6.88 µatm) and SHIP(smr) + FLOAT(PFZ) (RMSE = 8.0 µatm). This supports the hypothesis that resolving the intra-seasonal and seasonal variabilities of the meridional gradients is decisive in
735 minimizing uncertainties and bias in pCO_2 reconstructions. On the basis of this analysis, we propose that the optimal sampling scheme is the SHIP + nUSV because it provides not only a high temporal resolution (daily) of the large-scale meridional gradients but also combines speed to cover the required meridional spatial extent.

The nUSV Saildrones are still relatively new autonomous sampling platforms and their ability to withstand the stringent
740 weather and sea conditions in the Southern Ocean are still being assessed (Sutton et al., 2021). Recent deployments of Saildrones have been focused on zonal circumpolar tracks, which have been successful in proving the Saildrones as a robust sampling platform, and in observing the seasonal cycle of CO_2 fluxes in the sub-polar domain (Sutton et al., 2021). This approach is comparable to the zonal sampling of FLOATs (Fig. 1b) but with a higher temporal sampling frequency (daily vs 10-day). Notwithstanding the higher temporal sampling frequency from the Saildrone, the lack of a meridional spatial
745 component to the zonal sampling strategy limits its value in reducing the uncertainties and biases of any reconstructions that use them. Its inclusion in CO_2 flux reconstructions would improve the RMSE and mean bias error (MBE) relative to SOCAT-based reconstructions which, as discussed earlier, is not where autonomous sampling vehicles can add the best value (Tables 1 and 2).

750 Our work here shows that a zonal sampling strategy, while good for operational navigational reasons, is not the most efficient way to maximize the value of USV Saildrones sampling to resolve critical scales of variability necessary for a high confidence in the pCO_2 and inferred CO_2 flux reconstructions in the Southern Ocean. Furthermore, our study shows how by mixing the meridional sampling strategy (Lenton et al., 2006; Monteiro et al., 2010) with the current zonal sampling we can leverage the USV Saildrones to make sure we are not missing the meridional gradients.

755

4.5 Limitations of the study

In this study, our limitations were tied around three main points: the model used, the selected sub-domain, the existing shift in the seasonal cycle phasing of the model and data products, and the overfitting tendency of ML models. Here we discuss these limitations separately.

760

We only had one year of daily output of the high resolution coupled (NEMO-PISCES) model, BIOPERIANT12 (BP12). These BP12 model spatial (1/12º by 1/12º) and temporal (daily) resolutions influenced the designing of the observing system

https://doi.org/10.5194/bg-2021-344
Preprint. Discussion started: 3 January 2022

simulation experiments (OSSEs), therefore impacting the sampling approach of the synthetic platforms compared to their real-world counterparts. For example, unlike other sampling platforms that can be remotely piloted, FLOATs are a bit harder to simulate due to the way they operate. Thus, we could only mimic the 10-day sampling period and the deployment location and assume that they are randomly transported eastward by the water current. Since the Antarctic Circumpolar Current (ACC) moves from west to east the random walk we implemented is an adequate approximation and adds an element of stochasticity that is likely close to reality (Fig. 1b).

The selected Southern Ocean sub-domain combines regional and mesoscale gradients and features (such as eddies and fronts) that could challenge the reconstruction methods to better capture some variability scales such as the seasonal cycle of the meridional gradients (Fig. 8). However, the meridional gradients could be also associated with the meandering of the ACC fronts such as the SAF that overlaps with the sub-domain (Fig. 1). On the other hand, the assumption of the domain representativeness of the variability scales of the region could be a cause of concern as this would be applicable in regions where latitudinal gradients are strong. For example, the BP12 model output might not achieve this assumption based on a standard deviation of 9.1 µatm for the synthetic SHIP data compared to 20.96 µatm of the gridded shipboard observations (SOCAT) in the sub-domain.

Existing differences in the mean pCO_2 seasonal cycles of the model and data products (Fig. 2) could also result from processes that deterministic models such as the BIOPERIANT12 (NEMO-PISCES) cannot yet constrain due to a lack of understanding of the complete Southern Ocean carbonate system or mixed layer physics (Lenton et al., 2013; Monteiro et al., 2015; Mongwe et al., 2016). However, our knowledge of which one is right between model and data products remains limited.

Lastly, overfitting is a generally encountered issue in supervised machine learning (ML) problems. Although each of the two ML algorithms (FNN and GBM, see Sect. 2.4) we used was trained using best practices, the GBM algorithm encountered a bit more challenges with the overfitting compared to the FNN (cf. Table S3). While GBM have been proven to deal well with imbalanced or sparse datasets (Ke et al., 2017), it is more likely to overfit the training data because of the model potential for high complexity (Frery et al., 2017).

5 Conclusions

In this study we propose that in order to advance the uncertainties and biases from machine learning reconstructions in the Southern Ocean "beyond the wall", the seasonal cycle of the meridional gradient of pCO_2 needs to be resolved through a combination of high frequency (at least daily) observations spanning the meridional axis of the Southern Ocean. We showed that the reconstructed seasonal cycle anomaly and mean annual pCO_2 are highly sensitive to the seasonal sampling biases. The

https://doi.org/10.5194/bg-2021-344
Preprint. Discussion started: 3 January 2022

795 seasonal sampling bias comprises both the temporal and meridional spatial scales of variability. This may explain the significant winter positive bias in the reconstruction of the seasonal cycle of pCO_2 in the surface Southern Ocean, which may be contributing to an exaggeration of the winter outgassing CO_2 fluxes. This points to an urgent need to address the existing summer bias in the Southern Ocean SOCAT dataset and improve the present autonomous platform deployments, so they are better aligned to the integrated spatial and temporal sampling scale needs.

800

We confirmed that not resolving the high frequency (synoptic - sub-seasonal) variability results in insufficient decreases in mean biases and RMSEs for the reconstructed mean annual flux. Thus, we propose that present 10-day sampling periods from floats have a limited impact on uncertainties and biases because their temporal sampling scale does not resolve the intra-seasonal variability and their predominantly zonal Lagrangian sampling bias does not contribute sufficiently to resolving the 805 seasonal variability of the meridional gradient. If carbon float deployments are continued, we recommend increasing sampling frequency to 1 - 2 days particularly in high-EKE areas as well as a meridionally coherent deployment strategy that supports resolving the variability of the basin scale gradients. Wavegliders with higher temporal (daily) resolution in pseudo-mooring modes improve on floats (RMSEs, MBEs) because of the higher frequency but they still lack the meridional gradient reach. This study recommends that the use of Wavegliders in the reconstruction of CO_2 fluxes in the pseudo-mooring mode should 810 be discontinued and adopt a meridional dimension to the high temporal resolution (1-2 days). We showed that while the USV Saildrones in the present zonal sampling mode improve the RMSEs and biases, this is not the most efficient way to maximize their strengths of high sampling frequency and large spatial scale (by leveraging their speed). We propose that USV Saildrones are probably the optimal platforms to address the necessary integrated large-scale spatial and high-resolution temporal sampling, and thus have the most effective impact on the uncertainties and biases of the seasonal and mean annual 815 reconstruction of CO_2 fluxes in the Southern Ocean.

In summary, while ship-based observations (SOCAT) remain key to the reconstruction of CO_2 fluxes in the Southern Ocean through observations-based products, their summer seasonal sampling bias may be at the root of persistently elevated uncertainties and a positive winter seasonal bias in the reconstructions. This study recommends that the existing ship-based 820 approach to sampling surface ocean pCO_2 in the Southern Ocean be supplemented by year-round autonomous observations that resolve the seasonal cycle of the meridional gradients of surface ocean pCO_2.

Code and data availability

Supporting codes/scripts used for data analysis are contained in the GitHub repository https://github.com/Djeutsch/SOCCO-825 OSSE-v1. Data used in this study have been published in the online open-source repository Zenodo and can be accessed at https://doi.org/10.5281/zenodo.5788736 (Djeutchouang et al., 2021).

https://doi.org/10.5194/bg-2021-344
Preprint. Discussion started: 3 January 2022

Supplement

Author contributions

LMD is the lead author and developed the method and wrote the manuscript. LMD and PMSM conceived the study and performed the analysis. NC set up and ran the high resolution BIOPERIANT12 model used in the study. LG contributed to the development of the method and to editing the manuscript. MV contributed to the initial conceptualisation of the methods and proofread the manuscript. PMSM contributed substantially to the development of the manuscript and its reviews.

Competing interests

The authors have no competing interests.

Acknowledgements

This work is part of a PhD, and the study was supported by funding from the Department of Science and Innovation (DSI), South Africa, the National Research Foundation NRF-SANAP grants SNA170522231782, SNA170524232726 and CSIR Parliamentary Grant. We are grateful for the technical support and computational hours from the Centre for High Performance Computing (CSIR-CHPC). LG acknowledges funding from ETH Zürich and the European Commission through the COMFORT (grant no. 820989).

References

Aumont, O., Ethé, C., Tagliabue, A., Bopp, L., & Gehlen, M. (2015). PISCES-v2: An ocean biogeochemical model for carbon and ecosystem studies. *Geoscientific Model Development*, *8*(8), 2465–2513. https://doi.org/10.5194/gmd-8-2465-2015

Bakker, D. C. E., Pfeil, B., Olsen, A., Sabine, C. L., Metzl, N., Hankin, S., et al. (2012). Global data products help assess changes to ocean carbon sink. *Eos*, *93*(12), 125–126. https://doi.org/10.1029/2012EO120001

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-

https://doi.org/10.5194/bg-2021-344
Preprint. Discussion started: 3 January 2022

quality fCO2 data in version 3 of the Surface Ocean CO2 Atlas (SOCAT). *Earth System Science Data*, *8*(2), 383–413. https://doi.org/10.5194/ESSD-8-383-2016

Bushinsky, S. M., Landschützer, P., Rödenbeck, C., Gray, A. R., Baker, D., Mazloff, M. R., et al. (2019). Reassessing Southern Ocean Air-Sea CO2 Flux Estimates With the Addition of Biogeochemical Float Observations. *Global Biogeochemical Cycles*, *33*(11), 1370–1388. https://doi.org/10.1029/2019GB006176

Canadell, J. G., Monteiro, P. M. S., Costa, M. arco. H., Cunha, L. C. da, Cox, P. M., Eliseev, A. V., et al. (2021). Global Carbon and other Biogeochemical Cycles and Feedbacks. *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*.

Chapman, C. C., Lea, M. A., Meyer, A., Sallée, J. B., & Hindell, M. (2020). Defining Southern Ocean fronts and their influence on biological and physical processes in a changing climate. *Nature Climate Change*, *10*(3), 209–219. https://doi.org/10.1038/s41558-020-0705-4

Denvil-Sommer, A., Gehlen, M., Vrac, M., & Mejia, C. (2019). LSCE-FFNN-v1: A two-step neural network model for the reconstruction of surface ocean pCO2 over the global ocean. *Geoscientific Model Development*, *12*(5), 2091–2105. https://doi.org/10.5194/GMD-12-2091-2019

DeVries, T., Holzer, M., & Primeau, F. (2017). Recent increase in oceanic carbon uptake driven by weaker upper-ocean overturning. *Nature*, *542*(7640), 215–218. https://doi.org/10.1038/nature21068

Fay, A. R., & McKinley, G. A. (2013). Global trends in surface ocean p CO 2 from in situ data. *Global Biogeochemical Cycles*, *27*(2), 541–557. https://doi.org/10.1002/gbc.20051

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

Fay, Amanda R., Lovenduski, N. S., McKinley, G. A., Munro, D. R., Sweeney, C., Gray, A. R., et al. (2018). Utilizing the Drake Passage Time-series to understand variability and change in subpolar Southern Ocean pCO2. *Biogeosciences*, *15*(12), 3841–3855. https://doi.org/10.5194/bg-15-3841-2018

Fay, Amanda R., Gregor, L., Landschützer, P., McKinley, G. A., Gruber, N., Gehlen, M., et al. (2021). SeaFlux: Harmonization of air-sea CO2 fluxes from surface pCO2 data products using a standardized approach. *Earth System Science Data*, *13*(10), 4693–4710. https://doi.org/10.5194/ESSD-13-4693-2021

Frery, J., Habrard, A., Sebban, M., Caelen, O., & He-Guelton, L. (2017). Efficient Top Rank Optimization with Gradient Boosting for Supervised Anomaly Detection BT - Machine Learning and Knowledge Discovery in Databases. In M. Ceci, J. Hollmén, L. Todorovski, C. Vens, & S. Džeroski (Eds.) (pp. 20–35). Cham: Springer International Publishing.

Friedlingstein, P., Jones, M. W., Andrew, R. M., Bakker, C. E., Hauck, J., Le Quéré, C., et al. (2021). Global Carbon Budget 2021. *Earth System Science Data Discussions*, *2021*, 1–191. https://doi.org/10.5194/essd-2021-386

Gentemann, C. L., Scott, J. P., Mazzini, P. L. F., Pianca, C., Akella, S., Minnett, P. J., et al. (2020). Saildrone: Adaptively Sampling the Marine Environment. *Bulletin of the American Meteorological Society*, *101*(6), E744–E762. https://doi.org/10.1175/BAMS-D-19-0015.1

Gloege, L., McKinley, G. A., Landschützer, P., Fay, A. R., Frölicher, T. L., Fyfe, J. C., et al. (2021). Quantifying Errors in Observationally Based Estimates of Ocean Carbon Sink Variability. *Global Biogeochemical Cycles*, *35*(4), e2020GB006788. https://doi.org/10.1029/2020GB006788

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press. Retrieved from http://www.deeplearningbook.org

Grare, L., Statom, N. M., Pizzo, N., & Lenain, L. (2021). Instrumented Wave Gliders for Air-Sea Interaction and Upper Ocean Research. *Frontiers in Marine Science*, *8*(August), 1–21. https://doi.org/10.3389/fmars.2021.664728

Gray, A. R., Johnson, K. S., Bushinsky, S. M., Riser, S. C., Russell, J. L., Talley, L. D., et al. (2018). Autonomous Biogeochemical Floats Detect Significant Carbon Dioxide Outgassing in the High-Latitude Southern Ocean. *Geophysical Research Letters*, *45*(17), 9049–9057. https://doi.org/10.1029/2018GL078013

Gregor, L., & Gruber, N. (2021). OceanSODA-ETHZ: A global gridded data set of the surface ocean carbonate system for seasonal to decadal studies of ocean acidification. *Earth System Science Data*, *13*(2), 777–808. https://doi.org/10.5194/ESSD-13-777-2021

Gregor, L., Kok, S., & Monteiro, P. M. S. (2017). Empirical methods for the estimation of Southern Ocean CO2: Support vector and random forest regression. *Biogeosciences*, *14*(23), 5551–5569. https://doi.org/10.5194/bg-14-5551-2017

Gregor, L., Kok, S., & Monteiro, P. M. S. (2018). Interannual drivers of the seasonal cycle of CO₂ in the Southern Ocean. *Biogeosciences*, *15*(8), 2361–2378. https://doi.org/10.5194/bg-15-2361-2018

Gregor, L., Lebehot, A. D., Kok, S., & Scheel Monteiro, P. M. (2019). A comparative assessment of the uncertainties of global surface ocean CO2 estimates using a machine-learning ensemble (CSIR-ML6 version 2019a)-Have we hit the wall? *Geoscientific Model Development*, *12*(12), 5113–5136. https://doi.org/10.5194/gmd-12-5113-2019

Hauck, J., Völker, C., Wolf-Gladrow, D. A., Laufkötter, C., Vogt, M., Aumont, O., et al. (2015). On the Southern Ocean CO2 uptake and the role of the biological carbon pump in the 21st century. *Global Biogeochemical Cycles*, *29*(9), 1451–1470. https://doi.org/10.1002/2015GB005140

Hauck, Judith, Zeising, M., Le Quéré, C., Gruber, N., Bakker, D. C. E., Bopp, L., et al. (2020). Consistency and Challenges in the Ocean Carbon Sink Estimate for the Global Carbon Budget. *Frontiers in Marine Science*, *7*, 852. https://doi.org/10.3389/FMARS.2020.571720/BIBTEX

Hine, R., Willcox, S., Hine, G., & Richardson, T. (2009). The wave glider: A wave-powered autonomous marine vehicle. *MTS/IEEE Biloxi - Marine Technology for Our Future: Global and Local Challenges, OCEANS 2009*. https://doi.org/10.23919/OCEANS.2009.5422129

Holte, J., Talley, L. D., Gilson, J., & Roemmich, D. (2017). An Argo mixed layer climatology and database. *Geophysical Research Letters*, *44*(11), 5618–5626. https://doi.org/10.1002/2017GL073426

Iida, Y., Kojima, A., Takatani, Y., Nakano, T., Sugimoto, H., Midorikawa, T., & Ishii, M. (2015). Trends in pCO2 and sea–air CO2 flux over the global open oceans for the last two decades. *Journal of Oceanography*, *71*(6), 637–661.

https://doi.org/10.1007/S10872-015-0306-4/FIGURES/11

Jones, S. D., Le Quéré, C., Rödenbeck, C., Manning, A. C., & Olsen, A. (2015). A statistical gap-filling method to interpolate global monthly surface ocean carbon dioxide data. *Journal of Advances in Modeling Earth Systems*, *7*(4), 1554–1575. https://doi.org/10.1002/2014MS000416

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., et al. (2017). LightGBM: A highly efficient gradient boosting decision tree. *Advances in Neural Information Processing Systems*, *2017-Decem*(Nips), 3147–3155.

Keppler, L., & Landschützer, P. (2019). Regional Wind Variability Modulates the Southern Ocean Carbon Sink. *Scientific Reports*, *9*(1). https://doi.org/10.1038/S41598-019-43826-Y

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

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

Landschützer, Peter, Gruber, N., & Bakker, D. C. E. E. (2016). Decadal variations and trends of the global ocean carbon sink. *Global Biogeochemical Cycles*, *30*(10), 1396–1417. https://doi.org/10.1002/2015GB005359

Lenton, A, Tilbrook, B., Law, R., Bakker, D., Doney, S. C., & Gruber, N. (2013). Sea-air CO2 fluxes in the Southern Ocean for the period 1990–2009. *Biogeosciences Discussions*, *10*(1), 285–333. https://doi.org/10.5194/bgd-10-285-2013

Lenton, Andrew, Matear, R. J., & Tilbrook, B. (2006). Design of an observational strategy for quantifying the Southern Ocean uptake of CO2. *Global Biogeochemical Cycles*, *20*(4), 1–11. https://doi.org/10.1029/2005GB002620

Majkut, J. D., Carter, B. R., Frölicher, T. L., Dufour, C. O., Rodgers, K. B., & Sarmiento, J. L. (2014). An observing system simulation for Southern Ocean carbon dioxide uptake. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *372*(2019). https://doi.org/10.1098/RSTA.2013.0046

Maritorena, S., d'Andon, O. H. F., Mangin, A., & Siegel, D. A. (2010). Merged satellite ocean color data products using a bio-optical model: Characteristics, benefits and issues. *Remote Sensing of Environment*, *114*(8), 1791–1804. https://doi.org/10.1016/J.RSE.2010.04.002

McKinley, G. A., Fay, A. R., Eddebbar, Y. A., Gloege, L., & Lovenduski, N. S. (2020). External Forcing Explains Recent Decadal Variability of the Ocean Carbon Sink. *AGU Advances*, *1*(2), e2019AV000149. https://doi.org/10.1029/2019AV000149

Meinig, C., Lawrence-Slavas, N., Jenkins, R., & Tabisola, H. M. (2016). The use of Saildrones to examine spring conditions in the Bering Sea: Vehicle specification and mission performance. *OCEANS 2015 - MTS/IEEE Washington*. https://doi.org/10.23919/OCEANS.2015.7404348

https://doi.org/10.5194/bg-2021-344
Preprint. Discussion started: 3 January 2022

Meinig, C., Burger, E. F., Cohen, N., Cokelet, E. D., Cronin, M. F., Cross, J. N., et al. (2019). Public private partnerships to advance regional ocean observing capabilities: A saildrone and NOAA-PMEL case study and future considerations to expand to global scale observing. *Frontiers in Marine Science*, *6*(JUL), 1–15. https://doi.org/10.3389/fmars.2019.00448

Mongwe, N. P., Chang, N., & Monteiro, P. M. S. (2016). The seasonal cycle as a mode to diagnose biases in modelled CO2 fluxes in the Southern Ocean. *Ocean Modelling*, *106*, 90–103. https://doi.org/10.1016/j.ocemod.2016.09.006

Mongwe, N. P., Vichi, M., & Monteiro, P. M. S. (2018). The seasonal cycle of <i>p</i>CO₂ and CO₂ fluxes in the Southern Ocean: diagnosing anomalies in CMIP5 Earth system models. *Biogeosciences*, *15*(9), 2851–2872. https://doi.org/10.5194/bg-15-2851-2018

Monteiro, P. M. S., Monteiro, P. M. S., Monteiro, P. M. S., Monteiro, P. M. S., Monteiro, P. M. S., Monteiro, P. M. S., et al. (2010). A Global Sea Surface Carbon Observing System: Assessment of Changing Sea Surface CO2 and Air-Sea CO2 Fluxes, 702–714. https://doi.org/10.5270/OCEANOBS09.CWP.64

Monteiro, P. M. S. S., Gregor, L., Lévy, M., Maenner, S., Sabine, C. L., & Swart, S. (2015). Intraseasonal variability linked to sampling alias in air-sea CO2 fluxes in the Southern Ocean. *Geophysical Research Letters*, *42*(20), 8507–8514. https://doi.org/10.1002/2015GL066009

Munro, D. R., Lovenduski, N. S., Takahashi, T., Stephens, B. B., Newberger, T., & Sweeney, C. (2015). Recent evidence for a strengthening CO2sink in the Southern Ocean from carbonate system measurements in the Drake Passage (2002-2015). *Geophysical Research Letters*, *42*(18), 7623–7630. https://doi.org/10.1002/2015GL065194

Orsi, A. H., Whitworth, T., & Nowlin, W. D. (1995). On the meridional extent and fronts of the Antarctic Circumpolar Current. *Deep Sea Research Part I: Oceanographic Research Papers*, *42*(5), 641–673. https://doi.org/10.1016/0967-0637(95)00021-W

Pfeil, B., Olsen, A., Bakker, D. C. E., Hankin, S., Koyuk, H., Kozyr, A., et al. (2013). A uniform, quality controlled Surface Ocean CO2 Atlas (SOCAT). *Earth System Science Data*, *5*(1), 125–143. https://doi.org/10.5194/ESSD-5-125-2013

Le Quéré, C., Rödenbeck, C., Buitenhuis, E. T., Conway, T. J., Langenfelds, R., Gomez, A., et al. (2007). Saturation of the southern ocean CO2 sink due to recent climate change. *Science*, *316*(5832), 1735–1738. https://doi.org/10.1126/SCIENCE.1136188/SUPPL_FILE/PAP.PDF

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), 12,339-12,348. https://doi.org/10.1002/2017GL074837

Rödenbeck, C., Bakker, D. C. E., Gruber, N., Iida, Y., Jacobson, A. R., Jones, S., et al. (2015). Data-based estimates of the ocean carbon sink variability - First results of the Surface Ocean pCO2Mapping intercomparison (SOCOM). *Biogeosciences Discussions*, *12*(16), 14049–14104. https://doi.org/10.5194/bgd-12-14049-2015

Rödenbeck, Christian, Bakker, D. C. E., Metzl, N., Olsen, A., Sabine, C., Cassar, N., et al. (2014). Interannual sea–air CO_2 flux variability from an observation-driven ocean mixed-layer scheme. *Biogeosciences Discussions*, *11*(2), 3167–3207. https://doi.org/10.5194/bgd-11-3167-2014

https://doi.org/10.5194/bg-2021-344
Preprint. Discussion started: 3 January 2022

Sabine, C., Sutton, A., McCabe, K., Lawrence-Slavas, N., Alin, S., Feely, R., et al. (2020). Evaluation of a New Carbon Dioxide System for Autonomous Surface Vehicles. *Journal of Atmospheric and Oceanic Technology*, *37*(8), 1305–1317. https://doi.org/10.1175/JTECH-D-20-0010.1

Sabine, C. L., Hankin, S., Koyuk, H., Bakker, D. C. E., Pfeil, B., Olsen, A., et al. (2013). Surface Ocean CO2 Atlas (SOCAT) gridded data products. *Earth System Science Data*, *5*(1), 145–153. https://doi.org/10.5194/ESSD-5-145-2013

Stow, C. A., Jolliff, J., McGillicuddy, D. J., Doney, S. C., Allen, J. I., Friedrichs, M. A. M., et al. (2009). Skill assessment for coupled biological/physical models of marine systems. *Journal of Marine Systems*, *76*(1–2), 4–15. https://doi.org/10.1016/J.JMARSYS.2008.03.011

Sutton, A. J., Williams, N. L., & Tilbrook, B. (2021). Constraining Southern Ocean CO2 Flux Uncertainty Using Uncrewed Surface Vehicle Observations. *Geophysical Research Letters*, *48*(3), 1–9. https://doi.org/10.1029/2020GL091748

Takahashi, T., Olafsson, J., Goddard, J. G., Chipman, D. W., & Sutherland, S. C. (1993). Seasonal variation of CO2 and nutrients in the high-latitude surface oceans: A comparative study. *Global Biogeochemical Cycles*, *7*(4), 843–878. https://doi.org/10.1029/93GB02263

Takahashi, T., Sutherland, S. C., Wanninkhof, R., Sweeney, C., Feely, R. A., Chipman, D. W., et al. (2009). Climatological mean and decadal change in surface ocean pCO2, and net sea-air CO2 flux over the global oceans. *Deep-Sea Research Part II: Topical Studies in Oceanography*, *56*(8–10), 554–577. https://doi.org/10.1016/J.DSR2.2008.12.009

Takahashi, T., Sweeney, C., Hales, B., Chipman, D. W., Goddard, J. G., Newberger, T., et al. (2012). The changing carbon cycle in the southern ocean. *Oceanography*, *25*(3), 26–37. https://doi.org/10.5670/oceanog.2012.71

Talley, L. D., Rosso, I., Kamenkovich, I., Mazloff, M. R., Wang, J., Boss, E., et al. (2019). Southern Ocean Biogeochemical Float Deployment Strategy, With Example From the Greenwich Meridian Line (GO-SHIP A12). *Journal of Geophysical Research: Oceans*, *124*(1), 403–431. https://doi.org/10.1029/2018JC014059

Williams, N. L., Juranek, L. W., Feely, R. A., Johnson, K. S., Sarmiento, J. L., Talley, L. D., et al. (2017). Calculating surface ocean pCO2 from biogeochemical Argo floats equipped with pH: An uncertainty analysis. *Global Biogeochemical Cycles*, *31*(3), 591–604. https://doi.org/10.1002/2016GB005541

Wu, Y., Hain, M. P., Humphreys, M. P., Hartman, S., & Tyrrell, T. (2019). What drives the latitudinal gradient in open-ocean surface dissolved inorganic carbon concentration? *Biogeosciences*, *16*(13), 2661–2681. https://doi.org/10.5194/bg-16-2661-2019