UNIVERSITY OF CAPE TOWN
IYUNIVESITHI YASEKAPA • UNIVERSITEIT VAN KAAPSTAD

Sampling scale sensitivities in surface ocean pCO_2 reconstructions in the Southern Ocean

By

Laique Merlin Djeutchouang
(DJTMER001)

Submitted to the University of Cape Town in fulfilment of the requirements for the degree

Doctor of Philosophy

Department of Oceanography

FACULTY OF SCIENCES

Supervisor 1: Dr. Pedro M. S. Monteiro (Council for Scientific and Industrial Research)
Supervisor 2: Prof. Marcello Vichi (University of Cape Town)

May 26, 2023

Declaration

I hereby declare that the project proposal for this PhD dissertation was co-developed by myself and my PhD supervisors, where I played a major role in designing the focus of the project. The implementation of all the experiments, analyses, interpretation, presentation and writing of the dissertation represents my own work under the guidance of my PhD supervisors. The thesis was framed by myself, and the work contained here is my own except where explicitly stated in the text and has not been previously included or submitted for any other degree or qualification.

I confirm that I have been granted permission by the University of Cape Town's Doctoral Degrees Board to include the following publication(s) in my PhD thesis, and where co-authorships are involved, my co-authors have agreed that I may include the publication(s):

Djeutchouang, L. M., Chang, N., Gregor, L., Vichi, M., and Monteiro, P. M. S.: The sensitivity of pCO_2 reconstructions to sampling scales across a Southern Ocean sub-domain: a semi-idealized ocean sampling simulation approach, *Biogeosciences*, 19, 4171–4195, https://doi.org/10.5194/bg-19-4171-2022, 2022.

Acknowledgements

For you equipped me with strength for the battle and were the light for me to follow when the road was dark, I give you thanks Lord, my God, for everything in the name of Jesus Christ.

I would like to give thanks to Dr. Pedro Monteiro for his endless patience, careful guidance and understanding throughout this PhD. I really lack better words to thank you for the magnitude of the ocean CO_2 intellectual science space that you opened for me and walk me through. Similarly, I give special thanks to Prof. Marcello Vichi for his rigour and pragmatic insights. I really appreciate your mentorship in both doing and thinking about science.

I would like to thank Dr. Luke Gregor for introducing me to machine-learning techniques for carbon flux predictions. I learnt a lot from your mentorship. Thank you to Dr. Nicollete Chang for running the BIOPERIANT12-CNCRUN05A-S ocean model configuration and making the data available for use in this thesis.

For all their help in small and large ways, I thank the SOCCO group: Sandy, Sarah, Precious, Tommy, Nicolette, Isabelle, Pretty, etc. I also want to thank all SOCCO students for any kind of support you provide, particularly to Choaro and Asmita. This work was undertaken with financial support from the Department of Science and Innovation (DSI), the National Research Foundation (NRF) and the Council of Scientific and Industrial Research (CSIR). I would like to acknowledge the Centre for High-Performance Computing (CHPC) for providing the computational facilities and resources used for this PhD dissertation.

There are many friends and acquaintances who have helped to make this PhD journey a wonderful experience. I thank you all. I send a special thanks to Virginie, Romeo, Ricardo Harry, Rodrigue Franco, Atanas, Dieudonne, Karamael & Rachel Malisa, Mac Jugal, Mirriam, and Fabrice. Thank you to my Small Groups and Les Doc de Alvine for your spiritual support throughout this PhD. A special thanks to my mother and family. Thank you, Mom, you have been a constant source of care and support throughout this time.

Last in the list, but first in my heart, I thank Marthafuraha Nicholaus Massawe, for everything. You have been such a wonderful mother to our lovely daughter, Mildred Zuri Laique Djeutsch, through all the ups and downs of the past 4.5 years. Thank you.

And to you too, yes you who is reading this, thank you.

Abstract

The Southern Ocean plays a pre-eminent role in the global carbon-climate system. Model studies show that since the start of the preindustrial era, the region has absorbed about 75% of excess heat and 50% of the oceanic uptake and storage (42 ± 5 PgC) of anthropogenic CO_2 emissions. However, due to the spatial and seasonal sparseness of the Southern Ocean CO_2 observations (biased toward summer), this role is poorly understood. The seasonal sampling biases have hampered observation-based reconstructions of partial pressure of CO_2 at the surface ocean (pCO_2) using machine learning (ML) and contributed to the convergence of the root mean squared errors (RMSEs) of ML methods to a common limit known in the literature as the "wall". The hypothesis here is that addressing the critical missing sampling scale will get the community reconstructions of pCO_2 "over the wall". In this study, I explore the sensitivity of pCO_2 reconstructions to these observational scale gaps. Using a scale-sensitive sampling strategy means adopting a sampling strategy which addresses these observational limitations including intra-seasonal as well as seasonal sampling aliases in high eddy kinetic energy and mesoscale-intensive regions.

In increasing CO_2 sampling efforts in the Southern Ocean using autonomous sampling platforms such as floats, Wave Gliders and Saildrones, the community has tried to answer this problem, but the effectiveness of these efforts has not yet been tested. This study aims to do this evaluation and advance our understanding of the sampling scale sensitivities of surface ocean pCO_2 reconstructions from machine-learning techniques and contribute – through a scale-sensitive sampling strategy of observing platforms in the Southern Ocean – to breaking through the proverbial "wall". This aim was achieved through a series of observing system simulation experiments (OSSEs) applied to a forced mesoscale-resolving ($\pm$10km) ocean NEMO-PISCES physics-biogeochemistry model with daily output. In addition to underway ships, the sampling scales of the autonomous sampling platforms such as Floats, WaveGliders and Saildrones, on pCO_2 reconstructions were investigated in this series of OSSEs.

The primary results showed that two sampling scales, which Saildrones are able to address, are required to improve the RMSE scores of machine-learning techniques and then reduce uncertainties and biases in pCO_2 reconstructions. The two sampling scales include (1) the seasonal cycle of the meridional gradients and (2) the intra-seasonal variability. Based on the impacts of these two sampling scales on the RMSE scores and biases, it was found that resolving the seasonal cycle of the meridional gradient is the first-order requirement while resolving the intra-seasonal variability is the second. Applying the second-order requirement in the whole Southern Ocean to explore the sensitivity of the clustering choice to the two-step pCO_2 reconstruction (clustering- regression). It was found that using an ensemble of clustering methods in this two-step reconstruction performs far much better than using a clustering method. Using these findings, I proposed an observational strategy that is viable and strengthens the limitations in existing underway SOCAT

ship- and SOCCOM float-based reconstructions of surface ocean pCO_2. More specifically, I proposed a hybrid scale-sensitive sampling strategy for the whole Southern Ocean by integrating underway ships with Saildrones on winter lines.

The analysis of these multiple OSSEs indicates that improving the pCO_2 reconstructions requires scale-sensitive data to supplement the underway ship-based observations gridded in the SOCAT product. It was also found that scale-sensitive data consisting of high-resolution observations ($<$ 1 day) extending over the seasonal cycle and capturing the pCO_2 meridional gradients results in breaking through the proverbial "wall". These findings will contribute to an accurate mean annual global carbon budget which is critical for the trend of the ocean sink feedback on global warming as well as ocean acidification.

Acronyms

Acronyms	Descriptions
Chl-*a*	Chlorophyll-*a* concentration
CO_2	Carbon dioxide
CSIR	Council for Scientific and Industrial Research
ΔpCO_2	Gradient of pCO_2 between ocean and atmosphere with $\Delta pCO_2 = pCO_2^{ocn} - pCO_2^{atm}$
fCO_2	Furgacity of CO_2 in the gas phase (µatm)
FNN	Feed-forward Neural Network
GBM	Gradient Boosting Machine
MAE	Mean absolute error
MBE	Mean bias error
MG	Meridional gradient
ML	Machine Learning
MLD	Mixed layer depth
MOC	Meridional overturning circulation
NEMO	Nucleus for European Modelling Ocean
nUSV	New unmanned surface vehicle
OSSE	Observing systems simulation experiment
pCO_2	Partial pressure of CO_2 (µatm)
pCO_2^{atm}	partial pressure of CO_2 in the atmosphere
pCO_2^{ocn}	partial pressure of CO_2 at the surface ocean

PISCES	Pelagic Interactions Scheme for Carbon and Ecosystem Studies
PF	Polar Front
PFZ	Polar Frontal Zone
RMSE	Root mean squared error
SST	Sea surface temperature
SAF	Sub-Antarctic Front
SAM	Southern Annular Mode
SAZ	Sub-Antarctic Zone
SOCAT	Surface Ocean CO_2 ATlas
SOCCO	Southern Ocean Carbon-Climate Observatory
SOCCOM	Southern Ocean Carbon and Climate Observation Modelling
SOSCEx	Southern Ocean Seasonal Cycle Experiment
USV	Unmanned surface vehicle
WGM	Winter gap mitigation
$x\mathbf{CO_2}$	Dry mole fraction of atmospheric CO_2

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Statement of the problem

1.1.1 Study context

The primary driver of climate change is the additional human-induced component of atmospheric carbon dioxide (CO_2), which drives global warming and resulting climate change (Ciais et al., 2013; Canadell et al., 2021; Friedlingstein et al., 2021). The concentration of atmospheric CO_2 has increased over the last decades from approximately 340 ppm in 1980 to 412.4 ± 0.1 ppm in 2020 (Friedlingstein et al., 2022). This atmospheric CO_2 inflation is modulated by the ocean and land carbon sinks such that about 44% of airbone fraction of CO_2 (Canadell et al., 2021) due to anthropogenic CO_2 emission sources (such as land use change and fossil fuel combustion) remains in the atmosphere (Sabine et al., 2004; Friedlingstein et al., 2021). The remaining anthropogenic CO_2 ($\sim 60\%$) is then taken up into the terrestrial biosphere ($\sim 30\%$) (Ciais et al., 2013) and into the ocean ($\sim 26\%$ of total CO_2 emissions during the decade 2012-2021) (Friedlingstein et al., 2022).

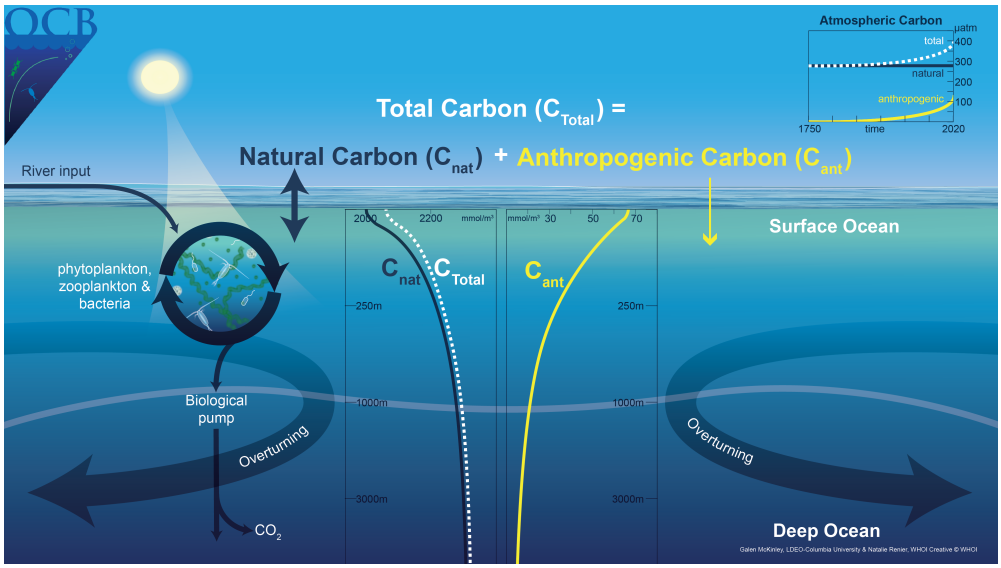


Figure 1.1: A new illustration on natural (C_{nat}) and anthropogenic (C_{ant}) components of the ocean carbon cycle, developed by the working group "Filling the gaps in observation-based estimates of air-sea carbon fluxes". Credit: Design: Natalie Renier, WHOI Creative ©WHOI; Concept: Galen McKinley, Columbia Univ., Lamont-Doherty Earth Observatory; Funding: Ocean Carbon & Biogeochemistry (OCB) Project Office (NSF, NASA).

The anthropogenic CO_2 uptake by the terrestrial biosphere is primarily due to a net increase in biomass,

whereas the anthropogenic carbon (C_{ant}) uptake by the ocean is mainly controlled by the physical carbon pump via the solubility of CO_2 in the seawater with subsequent transfer by water mass movement (cf. Figure 1.1). For the purpose of this thesis, our focus will be shifted toward the second mechanism.

The ocean is one of the major sinks of anthropogenic carbon dioxide (Ciais et al., 2013; Canadell et al., 2021). Over the historical period 1850-2020, the cumulated ocean carbon sink adds up to 170 ± 35 GtC, with two-thirds of this amount being taken up by the global ocean since 1960 (Khatiwala et al., 2013; Friedlingstein et al., 2021; Gruber et al., 2019, 2023) and other land-use activities (Ciais et al., 2013). More specifically, the ocean CO_2 sink has increased from 1.1 ± 0.04 GtC/yr in the 1960s to 2.8 ± 0.4 GtC/yr during 2011-2020 (Friedlingstein et al., 2021), with an interannual variation of the order of a few tenths of GtC/yr as shown in Figure 1.2. This assessment of anthropogenic CO_2 emissions and their redistribution (among the atmosphere, ocean and terrestrial biosphere) makes use the Global Ocean Biogeochemical Models (GOBMs)to derive the variability and trends in annual mean fluxes but now the reconstructions are used to evaluate the model trends (Hauck et al., 2020; Canadell et al., 2021). Specifically, observation based reconstructions are used for (1) calculating mean fluxes and (2) providing constraints for models, which provide confidence that the models can be used to derive trends over a longer period than that for which there are observations (Canadell et al., 2021).

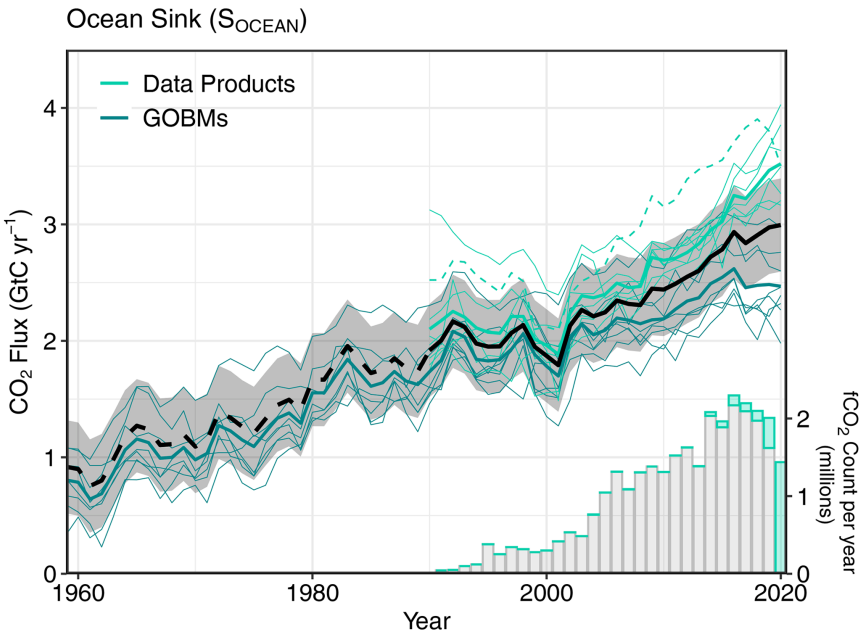


Figure 1.2: Ocean sinks (Friedlingstein et al., 2021): Comparison of anthropogenic air-sea CO_2 flux showing the budget values of ocean sink (black; with uncertainties in grey shading), data-product-based values (cyan; with Watson et al. (2020) in dashed line as not used for ensemble mean), and individual ocean-model-based values (teal). Data products are adjusted for the pre-industrial ocean source of CO_2 from river input to the ocean with the average of 0.45 ± 0.18GtC/yr by Jacobson et al. (2007), and 0.78 ± 0.41GtC/yr by Resplandy et al. (2018)

According to the 2019 assessment of the Global Carbon Project, the ocean took up on average approximately 2.5 ± 0.6 PgC/yr or $23 \pm 5\%$ of the total anthropogenic emissions of CO_2 over the decade 2009-2018 (Friedlingstein et al., 2021; Hauck et al., 2020). As longer anthropogenic activities such as cement manufac-

turing and fossil-fuel burning keep on increasing, the levels of atmospheric CO_2 emissions will continue to rise as a consequence (Friedlingstein et al., 2021). This increase in atmospheric CO_2 is responsible for the increases in the global average temperatures (Prentice et al., 2001) because the CO_2 in the atmosphere acts as a powerful greenhouse gas. Oceanic uptake of CO_2 slows the rate of atmospheric CO_2 increase and thus highlights the pivotal role the ocean plays in slowing the rate of climate change (Lenton et al., 2006). For atmospheric CO_2 to get into the ocean, the primary pathway is through the gas exchange at the atmosphere-ocean boundary. This means the ocean takes up CO_2 primarily through the exchange of CO_2 across the air-sea interface, which is a function of the pCO_2 difference between the surface ocean pCO_2 (pCO_2^{ocn}) and the atmospheric pCO_2 (pCO_2^{atm}). This is expressed as follows:

$$\Delta pCO_2 = pCO_2^{ocn} - pCO_2^{atm}. \tag{1.1.1}$$

The growth rate of oceanic pCO_2 relative to the atmospheric pCO_2 provides information on the evolution of the strength of the sink or source of atmospheric CO_2 in time (Lenton et al., 2013). However, the direction of the air-sea CO_2 flux is defined by ΔpCO_2 across the atmosphere-ocean interface with additional controls from the gas transfer velocity and CO_2 solubility setting the magnitude. Furthermore, the metric used for assessing the ocean's role in CO_2 emissions is tightly linked to the air-sea CO_2 flux, F_{CO_2}, that is estimated from the bulk formulation established after Wanninkhof (1992).

$$F_{CO_2} = k_W \times k_0 \times \Delta pCO_2 \tag{1.1.2}$$

In this bulk formulation (Equation 1.1.2), the first parameter k_W describes the wind-driven [kinetic] gas transfer velocity of CO_2 between the ocean and the atmosphere. It is approximated as a quadratic function of wind speed (U_{10}) at ~ 10 m of sea surface (Sweeney et al., 2007; Wanninkhof, 2014). The second parameter k_0 is a temperature and salinity-dependent function describing the aqueous-phase solubility of CO_2 in the seawater. Lastly, ΔpCO_2 represents the gradient of pCO_2 at the interface between the ocean and atmosphere (see Equation 1.1.1). The basic assumption associated with this bulk aerodynamic metric (Equation 1.1.2) is that the CO_2 sink is determined by the pCO_2 gradient ΔpCO_2 and the CO_2 flux F_{CO_2} is described by the gas transfer velocity (k_W) with the wind speed (U_{10}) as the principal driving-kinetic parameter (Wanninkhof, 1992, 2014). When there is a difference between pCO_2^{ocn} and pCO_2^{atm}, the ocean takes up or releases CO_2. This means that the direction of the flux, F_{CO_2}, is determined by the sign of ΔpCO_2. By convention, negative values of F_{CO_2} (i.e., $\Delta pCO_2 < 0$) correspond to transport from the atmosphere to the ocean, whereas positive values of F_{CO_2} (i.e., $\Delta pCO_2 > 0$) represents transport from the ocean to the atmosphere.

Recent years have seen an international collaborative effort that has assembled high quality-controlled data of surface ocean F_{CO_2} observations, the Surface Ocean F_{CO_2} Atlas (SOCAT) (Sabine et al., 2013; Bakker et al., 2016). This database has played a pivotal role in most recent studies focusing on reconstructing

the surface ocean pCO_2 from observations and subsequently inferring F_{CO_2} (e.g., Rödenbeck et al., 2015; Landschützer et al., 2016; Gregor et al., 2019; Denvil-Sommer et al., 2019). However, most observations contributing to the SOCAT database are obtained from underway sampling platforms on board volunteer observing ships that are constrained in space and time (Bakker et al., 2016; Gregor et al., 2019).

Observation-based CO_2 products thoroughly rely on the interpolation of pCO_2 at the surface ocean, which is derived from the underway SOCAT ship measurements of the surface ocean CO_2 fugacity, and ancillary variables from satellite and reanalysis oceanic data. This interpolation or mapping is made up of a variety of methods ranging from novel statistical inference (e.g., Rödenbeck et al., 2014; Watson et al., 2020), to machine learning (e.g., Landschützer et al., 2013; Zeng et al., 2015; Gregor et al., 2017; Denvil-Sommer et al., 2019). All these novel statistical and machine-learning methods provide estimates of the global ocean carbon sink and its variability at seasonal and inter-annual timescales (Rödenbeck et al., 2015; Gregor et al., 2019).

Accurate estimation of F_{CO_2} matters because most of the CO_2 data products resulting from some of the estimation methods including Landschützer et al. (2016) and Gregor et al. (2019) are essential for an accurate mean annual global carbon budget even though they do not capture the mean decadal (Gloege et al., 2021). The global carbon budget is an estimate of the flows of CO_2 bwtween the major reservoirs of Earth's carbon cycle, including the atmosphere, land, and ocean (Friedlingstein et al., 2021; Canadell et al., 2021). In other words, this is a way of accounting for the sources (emissions) and sinks (absorption) of CO_2 in the Earth system (Friedlingstein et al., 2022). On the other hand, accurate estimation of ocean CO_2 fluxes is critical for approximating the trend of the ocean CO_2 sink feedback on global warming as well as ocean acidification (Canadell et al., 2021). However, these methods for estimating the ocean F_{CO_2} fluxes exclusively rely on sparse and limited observations predominantly from underway ship sampling which is seasonally biased (Bushinsky et al., 2019; Gregor et al., 2019). This, therefore, hampers the reliability of these observation-based estimates, particularly in data-sparse regions such as the Southern Ocean and in winter when these regions are more inaccessible.

1.1.2 Definition of the problem in the Southern Ocean

The spatiotemporal density of surface ocean CO_2 *in-situ* observations is not homogeneous within the Southern Hemisphere (Gregor et al., 2019; Gloege et al., 2021), particularly in the Southern Ocean that is poorly sampled in both space and time (Lenton et al., 2006; Monteiro et al., 2010; Fay et al., 2018; Bushinsky et al., 2019) as one can see in Figure 1.3 (Fay et al., 2018). Therefore, sparse data coverage and the lack of observations covering the full seasonal cycle (see Figure 1.3) are thus major sources of uncertainties and biases in observation-based estimates of CO_2 sink in the Southern Ocean. These spatiotemporal limitations challenge empirical mapping methods and models to capture the whole-scale variability of pCO_2 (Gregor

et al., 2019), which may result in unknown impacts of reconstruction biases and unceratinties, and hence restricts our reliability on observation-based estimates (Rödenbeck et al., 2015; Denvil-Sommer et al., 2019). These observational limitations have been a significant cause of the convergence of the pCO_2 reconstruction performances to a relatively common value – proverbially known as the "hitting the wall" or simply the "wall" (Gregor et al., 2019).

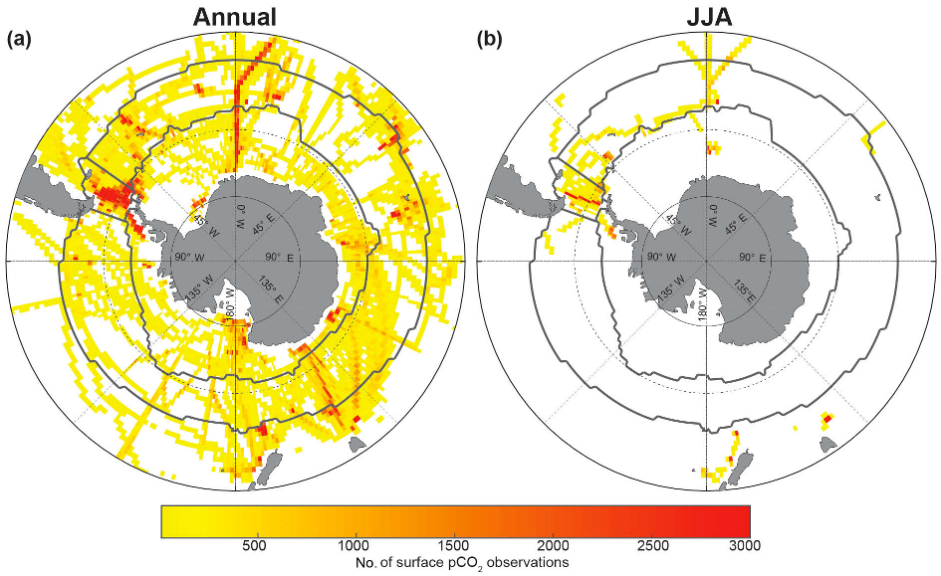


Figure 1.3: The density of observations of the Southern Ocean surface CO_2 in the SOCAT v5 (released in 2017) with each $1^\circ \times 1^\circ$ grid cell restricted to years 2002-2016 (Fay et al., 2018; Gray et al., 2018) where **(a)** and **(b)** respectively represent all months of the year and winter (June, July and August, JJA).

In the Southern Ocean, existing quality controlled and gridded observations (SOCAT) used in the reconstructions of pCO_2^{ocn} are summer and spatially biased as they are primarily from volunteering underway ships (Bakker et al., 2016). In the illustrations of the ship sampling density (Figures 1.3 and 1.4), these seasonal and spatial observational biases in the region become clearly apparent. Despite this data sparseness in the region, researchers have developed and maintain global monthly gridded observation-based products such as Jena-MLS (Rödenbeck et al., 2015), MPI-SOMFFN (Landschützer et al., 2016), and CSIR-ML6 (Gregor et al., 2019), for global surface ocean pCO_2^{ocn} and air-sea CO_2 fluxes.

In an increasing effort to respond to these sampling limitations and fill the spatiotemporal observational gaps in the Southern Ocean, the carbon cycle community has been trying for the past few years to increase the coverage of surface ocean CO_2 observations in the Southern Ocean through autonomous observing platforms such as biogeochemical profiling floats (Williams et al., 2017; Gray et al., 2018; Bushinsky et al., 2019), Wave Gliders (Monteiro et al., 2010; Nicholson et al., 2022), and Saildrones (Sutton et al., 2021). Particularly, the project called Southern Ocean Carbon and Climate Observation and Modelling (SOCCOM) (https://soccom.princeton.edu/) started with the aim to deploy about 200 biogeochemical profiling floats in the region over a period of 5 years between 2015 to 2020 (Williams et al., 2017; Fay et al., 2018). Figure 1.4 illustrates some of the results of this effort with increased data coverage in space and time in the Southern

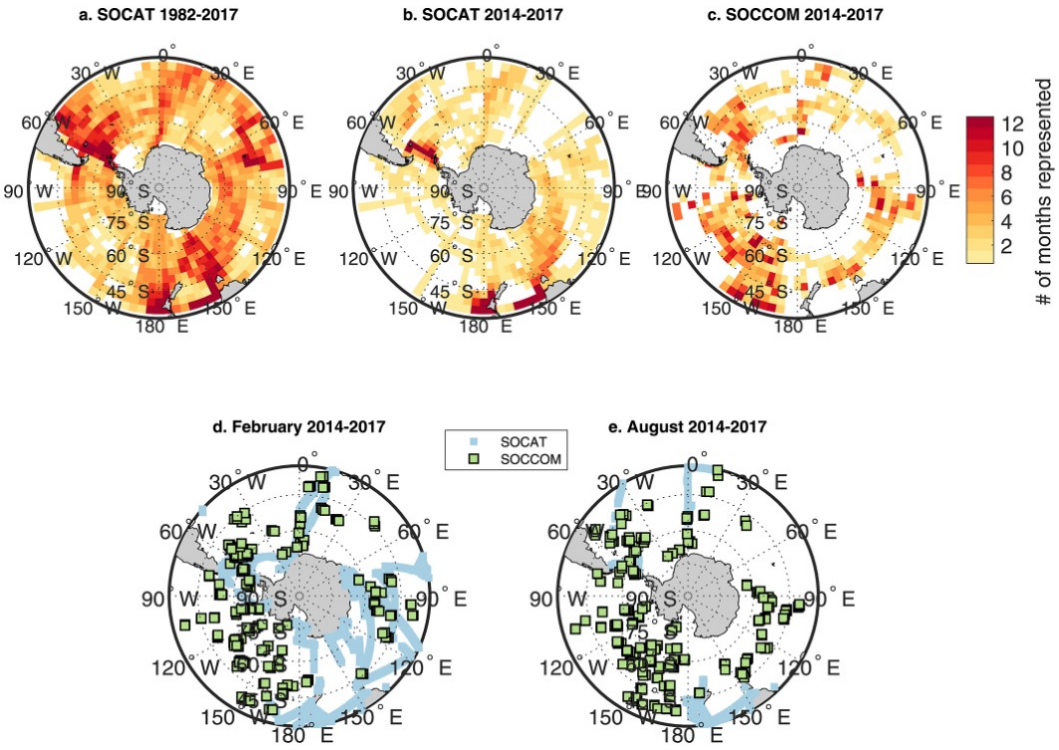


Figure 1.4: Depictions of the temporal and spatial distributions of ship-(SOCAT) and float-(SOCCOM) based samplings from Bushinsky et al. (2019) in the Southern Ocean (40°S). Panels **(a)** and **(b)** illustrate the sampling density for the SOCAT data within the periods 1982-2017 and 2014-2017, respectively; panel **(c)** illustrates the sampling density for SOCCOM float data within the period 2014-2017; and the colors represent the number of calendar months (maximum 12) with any data present in a grid cell of 3°-latitude by 4°-longitude. Panels **(d)** and **(e)** respectively illustrating the sampling density for summer (February) and winter (August) months for both SOCAT and SOCCOM data within the period 2014-2017, also show the bias in shipboard observations toward warming months as well as remaining gaps in SOCCOM float coverage (see Bushinsky et al. (2019) for more details).

Ocean over this 5-year period. The SOCCOM floats are different from traditional floats Argo floats in the sense that some include ice avoidance software, while all carry some combination of additional biogeochemical sensors for pH, oxygen, fluorescence, nitrate, and backscattering (Carter et al., 2016; Williams et al., 2017). Other contributions to this sampling effort have been done using Wave Gliders through the Southern Ocean Seasonal Cycle Experiment (SOSCEx) project (Monteiro et al., 2010; Nicholson et al., 2022).

The seasonal sampling biases have hampered observation-based reconstructions of the surface ocean pCO_2 using machine learning (ML) and contributed to the convergence of the performance of existing ML methods, in terms of root mean squared errors (RMSEs), to a common limit known in the literature as the "wall". Gregor et al. (2019) proposed that scale-sensitive integrated multi-platform sampling of surface ocean pCO_2 in high-latitude regions (such as in the Southern Ocean) should be the top priority for the carbon cycle community and suggested that optimized simulation sampling experiments should be used to understand the spatial and temporal requirements of pCO_2 in these regions and periods. The study focuses on these points by investigating how sensitive pCO_2 reconstructions are to sampling scales of CO_2 observing platforms and used the RMSEs of the reconstruction approaches as the main assessment metric of the uncertainties. Thus, the hypothesis here is that addressing the critical missing sampling scales will get the

community reconstructions of pCO_2 "over the wall".

1.1.3 Differences in CO_2 products in the Southern Ocean

All empirical methods involved in the estimation of the surface ocean pCO_2 are not yet optimal for making good predictions (Keppler and Landschützer, 2019). These methods are just gap-filling approaches; they are thus more appropriate for interpolations in the study domain. Data sampling limitations challenge mapping methods and models to capture the whole scale variability of pCO_2^{ocn} (Gregor et al., 2019), and hence restrict our reliability on observation-based estimates (Rödenbeck et al., 2015; Denvil-Sommer et al., 2019).

Furthermore, these limitations result in fluctuated reconstructions with significant uncertainties and biases in pCO_2^{ocn} estimates (Denvil-Sommer et al., 2019), and then large discrepancies between estimates from different methods and data products (Hauck et al., 2020; Friedlingstein et al., 2021). Notwithstanding the increasing evidence for the importance of the Southern Ocean CO_2 uptake to global and regional climate change, particularly its strengthening as a carbon sink in the 2000s following the weakening in the 1990s (Landschützer et al., 2015; Frölicher et al., 2015; Gregor et al., 2019), significant differences are found in various data products in the region (Williams et al., 2017; Gray et al., 2018). For example, Figure 1.5 illustrates differences in the pCO_2^{ocn} estimates throughout the Southern Ocean (south of 40°S) from two widely used data products SOM-FNN and CSIR-ML6. In Figure 1.5b, it can be seen that larger differences in the magnitude and phasing of pCO_2^{ocn} from the two data products exist between 1982-1998. However, from 1998 only the difference in phasing got improved until about 2012, and then the differences in both magnitudes and phasing started getting lower and lower.

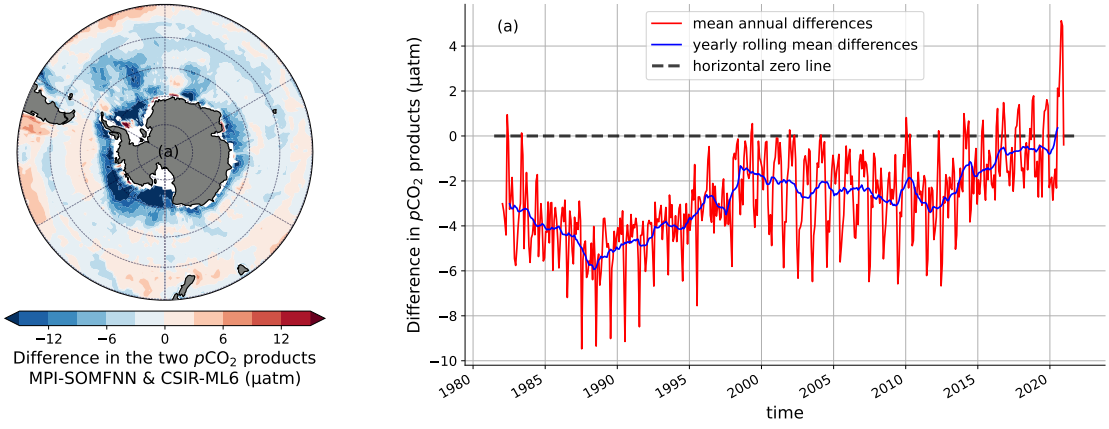


Figure 1.5: Differences in SOM-FNN (Landschützer et al., 2016) and CSIR-ML6 (Gregor et al., 2019) data products of surface ocean pCO_2 (μatm) in the Southern Ocean south of 40°S. Panel **(a)** shows the difference map while panel **(b)** shows the inter-annual differences (1982-2020). The dashed line (black) corresponds to the zero line.

1.2 Why the Southern Ocean?

The Southern Ocean can be defined as the ocean that extends from the landmass of the Antarctic continental block to about 30-40°S or the more dynamical boundary of the sub-Antarctic front (STF) and connects the Atlantic, Indian and Pacific Oceans (Rintoul et al., 2001; Orsi et al., 1995). The term Southern Ocean was first used to describe this ocean area by the famous mariner James Cook[1]. Furthermore, the ocean sink of anthropogenic carbon is not equally distributed throughout the global ocean (Sarmiento and Gruber, 2006; Gregor et al., 2018). It is the strongest in high-latitude areas such as the Southern Ocean, home to deep and intermediate water formations (Gregor et al., 2019; Denvil-Sommer et al., 2019).

1.2.1 Southern Ocean sensitivity to climate

In recent years, studies that started resolving the interannual variability of air-sea CO_2 fluxes revealed the climate sensitivity of the Southern Ocean (Landschützer et al., 2015). According to McNeil and Matear (2008), the uptake of anthropogenic CO_2 happens on top of the natural CO_2 outgassing, which is assumed to have been stable since the pre-industrial period (Friedlingstein et al., 2021).

Over the last century, the Southern Ocean has played a role as the largest sink of excess heat (75%) and about 50% of the ocean uptake and storage (42±5 PgC) of anthropogenic CO2 emissions since the start of the industrial era (Sabine et al., 2004; Frölicher et al., 2015; Gregor et al., 2019). Thus, the magnitude of carbon fluxes in the Southern Ocean is a balance between the outgassing of natural CO_2 and the uptake of anthropogenic CO_2, making the region has important implications on the global carbon cycle and climate (Takahashi et al., 2009; Lenton et al., 2013). This means that to better understand the global carbon cycle as well as improve the international effort to mitigate climate change and improve projections, it is critical to improve the accuracy of the Southern Ocean CO_2 sink estimates.

The present state of the Southern Ocean CO_2 uptake is assumed to be variable, which suggests an imbalance between the outgassing of natural CO_2 and the uptake of anthropogenic CO_2 (DeVries et al., 2017). The modulation of the air-sea CO_2 exchange by the Southern Ocean has been the focus of many studies (LeQuéré et al., 2007; Landschützer et al., 2015; Frölicher et al., 2015; DeVries et al., 2017; Gruber et al., 2019; Gregor et al., 2018; McKinley et al., 2020). In their study, the authors found that the total rate of Southern Ocean uptake of CO_2 weakened in the decade 1990s, but the region regained its strength to take up CO2 in the next decade (2000s) beginning in 2002 until at least 2012 (Figure 1.6). In the 1990s, there was increased outgassing of natural CO_2 resulting from the weakening of the Southern Ocean carbon sink (LeQuéré et al., 2007), which was due to a southerly shift in westerly winds (Keppler and Landschützer, 2019). On the other hand, Landschützer et al. (2015) reported a reinvigoration of CO_2 sink in the South-

[1] James Cook (1728 - 1779) was a British explorer, mariner, cartographer, and captain in the British Royal Navy.

ern Ocean in the 2000s as a result of a change in the zonal nature of atmospheric circulation that led to a basin-specific interaction between carbonate chemistry in the ocean and the thermal response of CO_2 partial pressure (pCO_2). This shows the complexity of CO_2 response to drivers. The authors used *in-situ* observations and a two-step machine-learning method to investigate the decadal evolution of carbon sink in the Southern Ocean.

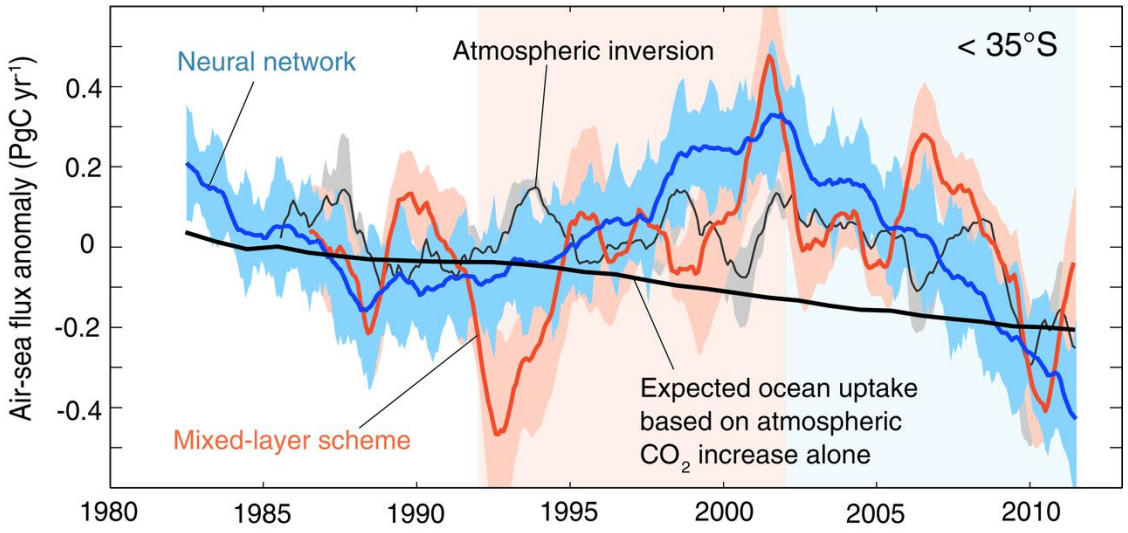


Figure 1.6: Southern Ocean (south of 35°S) CO_2 sink anomalies (Landschützer et al., 2015).

Based on recent studies, there are currently three main hypotheses to explain the decadal variations in the Southern Ocean CO_2 sink. These hypotheses include (1) the develpment of Zonal Wavenumber 3 pattern (Keppler and Landschützer, 2019; Gruber et al., 2019) which is potentially linked to stratospheric Ozone, (2) changes in the shallow overturning circulation (DeVries et al., 2017), and (3) volcanic activity and external forcing (McKinley et al., 2020). They contribute in strengthen our understanding on the Southern Ocean's sensitivity to climate. Furthermore, an additional fourth hypothesis related to increased storms in the region may be considered (Gregor et al., 2018), however that could potentially be linked with the first hypothesis, concerning the development of the Zonal Wavenumber 3 large-scale pattern.

Two major oceanic mechanisms can modulate the varaibility of the air-sea CO_2 flux or pCO_2 gradient growth rates to be either faster than the atmospheric CO_2 growth rates, leading to a decreasing sink of atmospheric CO_2, or slower than atmospheric CO_2 , leading to an increasing sink of atmospheric CO_2 (Equations 1.1.1 and 1.1.2). The first major mechanism, called bilogical pump, is the changes in the biological primary production (Sarmiento and Gruber, 2006; Landschützer et al., 2015; Gregor et al., 2018). This mechanism involves the transfer of carbon within the ocean's ecosystems and plays a significant role in the long-term storage of carbon. In the biological pump, carbon is taken up by marine organisms, particularly phytoplankton, through the process of photosynthesis. Phytoplankton use CO_2 and sunlight to produce organic matter (Sarmiento and Gruber, 2006).

The second mechanisms, called solubility pump, corresponds to changes in ocean physics that can lead to

changes in sea surface temperature, sea surface salinity and alkalinity, which affect the solubility of dissolved inorganic carbon (DIC) in the upper ocean (Sarmiento and Gruber, 2006; Landschützer et al., 2015; Gruber et al., 2019). More specifically, when atmospheric CO_2 levels increase due to human activities, such as the burning of fossil fuels, a portion of that excess CO_2 is absorbed by the ocean (Friedlingstein et al., 2022). This absorption process occurs through the solubility cycle, where CO_2 dissolves in seawater and forms carbonic acid, which then dissociates into bicarbonate and carbonate ions. The dissolved CO2 is effectively "pumped" into the ocean, reducing the amount of CO_2 in the atmosphere (Sarmiento and Gruber, 2006) and mitigating the greenhouse effect that contributes to global warming (Friedlingstein et al., 2022).

In the Southern Ocean, observation-based estimates of contemporary air-sea CO_2 flux typically find substantial net uptake with climatological values of -0.8 and -1.0 PgC/yr (in the 2000s) respectively estimated by Takahashi et al. (2009) and Landschützer et al. (2015), where negative sign denotes flux into the ocean. Furthermore, on longer time scales, changes in this region's carbon flux have also been implicated in modulating glacial-interglacial cycles (Sigman et al., 2010), and significant variability exists on interannual to decadal time scales (LeQuéré et al., 2007; Lenton et al., 2013; Landschützer et al., 2015). Based on recent estimates from global climate model simulations, a modern Southern Ocean CO_2 uptake was in the range of -0.5 and -1.3 PgC/yr in 2015 (Nevison et al., 2016).

1.2.2 Southern Ocean circulation and CO_2 sink

The Earth can be seen as an ocean planet due to $\sim$71% of its surface coverage by water and up to $\sim$6% of sea ice (Grassl, 2001). Therefore, the uptake and storage of carbon and heat are dominated by the oceans. Large-scale oceanic flows in the open ocean are nearly in geostrophic and hydrostatic balance (Rintoul et al., 2001; Marshall and Speer, 2012).

The Southern Ocean plays a unique role in connecting all the Earth's Oceans (Figure 1.7) and influencing the Earth's climate system. In addition, the vital role the region has in the global anthropogenic CO_2 uptake is dominantly due to its unique circulation (Marshall, 2003; Marshall and Speer, 2012). The Southern Ocean is geographically opened at the Drake Passage and carries the Antarctic Circumpolar Current (ACC) which flows continuously eastward around the Antarctic continent as shown in Figure 1.7.

A strongly divergent surface flow is driven by the westerly winds and allows old Upper Circumpolar Deep Water that carries little anthropogenic CO2 to come back to the surface ocean at the Southern Boundary (SB) or Polar Front (Carter et al., 2008; Marshall and Speer, 2012; Talley, 2013; Terhaar et al., 2021) that is the southern limit of the ACC (Figure 1.7). When a small fraction of upwelled and surface water moves southward and is converted into Antarctic Bottom Waters (Talley, 2013; Terhaar et al., 2021), the largest fraction flows northward through Ekman transport while taking up large amounts of anthropogenic CO_2 via exchange of gas between atmosphere and ocean (DeVries et al., 2017; Terhaar et al., 2021). Ultimately, these

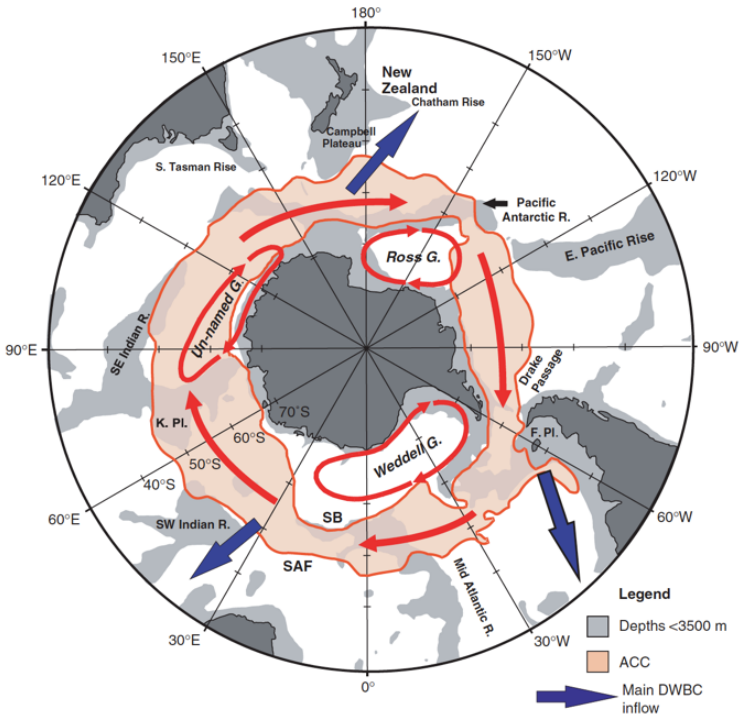


Figure 1.7: An illustration – after Carter et al. (2008) – of a major circulation feature of the Southern Ocean: the Antarctic Circumpolar Current (ACC) and its links to the meridional overturning circulation (MOC). The map describes the Southern Ocean regions with inflow and outflow connections with the southern basin of the Atlantic, and Indian Pacific Oceans. Specifically, it shows (1) the ACC constrained by the Sub-Antarctic Front (SAF) and the Southern Boundary (SB) or the southern limit of Upper Circumpolar Deep Water (UCDW); (2) the Ross, Weddell, and Un-named subpolar gyres (G.); (3) the main exit points of Deep Western Boundary Current (DWBC) from the Southern Ocean (blue arrows). Lastly, are shown the main annotations of the bathymetry elevations F. Pl., K. Pl., and R. corresponding to Falkland Plateau, Kerguelen Plateau and Ridge, respectively.

northward-flowing water masses are transformed into Sub-Antarctic Mode Water (SAMW) (Talley, 2013) and Antarctic Intermediate Water (AIW) by means of surface heat uptake and mixing with southward flowing water from mid-latitude regions (Gruber et al., 2019; Terhaar et al., 2021). Now, enriched in anthropogenic carbon, SAMW and AIW are then subducted in the vicinity of the Sub-Tropical Front (STF) below the light Sub-Tropical waters into the deeper ocean (Talley, 2013; Terhaar et al., 2021). The formation rate of SAMW and AIW primarily dictate the amount of anthropogenic carbon taken up in the Southern Ocean in such a way that other factors such as the buffer capacity and the air-sea equilibration time of pCO_2 become less important (Caldeira and Duffy, 2000; Sabine et al., 2004; DeVries et al., 2017; Terhaar et al., 2021).

The Southern Ocean opening at the Drake Passage creates a mixture of Sverdrup and non-Sverdrup balance for the general circulation, hence the Antarctic subpolar gyres are observed, for example, at Weddell and Ross Seas (Carter et al., 2008). It is within the latitude band of the Drake Passage (absence of land barriers) that the strong eastward flow of the ACC connects each of the three basins corresponding to the Atlantic, Indian and Pacific Oceans (Rintoul et al., 2001; Rintoul, 2008; Carter et al., 2008). The ACC, as it encircles Antarctica, is the world's largest ocean current and the most dominant circulation feature of the Southern Ocean (Rintoul, 2008). As illustrated in Figure 1.7, the ACC circulation includes the Amundsen

Sea, Bellingshausen Sea, Ross Sea, Weddell Sea, a major part of the Drake Passage, and a small part of the Scotia Sea (Carter et al., 2008; Metzl et al., 2006; Lenton et al., 2013). This ACC circulation is particularly unique as it combines the upwelling of deep circumpolar waters, very rich in carbon and nutrients, and the subduction of surface freshwater (Gruber et al., 2009; Sallée et al., 2013; Gruber et al., 2019). The ACC circulation allows the Southern Ocean to close the overturning circulation of the global ocean (Marshall and Speer, 2012; DeVries et al., 2017) while maintaining the nutrient balance of low-latitude oceans (Sarmiento et al., 2004) and sequestering anthropogenic CO_2 and excess heat from the atmosphere (Frölicher et al., 2015).

1.3 OSSEs and empirical modelling of pCO_2

1.3.1 The current state of OSSEs

Since the 1980s, Observing System Simulation Experiments (OSSEs) have been developed and conducted for extensive numerical weather prediction to explore the impact of new observational data on analysis quality and forecast skill (Zeng et al., 2020). They were first developed at the Goddard Space Flight Center of the National Aeronautics and Space Administration (NASA), and at the Atlantic Oceanographic and Meteorological Laboratory of the National Oceanic and Atmospheric Administration (NOAA), in collaboration with academic partners, private enterprises, and operational data assimilation centers (Xue et al., 2006; Atlas, 1997; Zeng et al., 2020). These OSSEs have been used to correctly quantify the potential for several proposed satellite observing systems to improve weather analysis and prediction prior to their launch. Nowadays, the adoption of OSSEs in different research sectors has improved quantitative and objective assessment capabilities to evaluate operational and future observation system impacts and trade-offs (Xue et al., 2006; Halliwell et al., 2014; Atlas, 1997; Kamenkovich et al., 2017).

NASA has broadened the definition of OSSE to include two sets of approaches: sampling OSSE and retrieval OSSE. The retrieval OSSE approach seeks to quantify the degree to which prospective observations provide information on the geophysical parameter or variable of interest, whereas a sampling OSSE is used to figure out whether a chosen observing system candidate has sufficient spatial and temporal sampling coverage to address a specific science question (Atlas et al., 2015; Zeng et al., 2020). However, in addition to assessing the observation sufficiency, outcomes from the retrieval OSSE can be used to further analyse uncertainties and biases in a predictive OSSE framework. Both sets of OSSE approaches are schematically depicted in Figure 1.8.

The recent literature points to two key types of OSSEs: the traditional global forecasting OSSE which has reached the most advanced stage with extensive peer-reviewed papers, and the historical or sampling OSSE approach which is relatively new and needs further development (Atlas et al., 2015; Zeng et al., 2020).

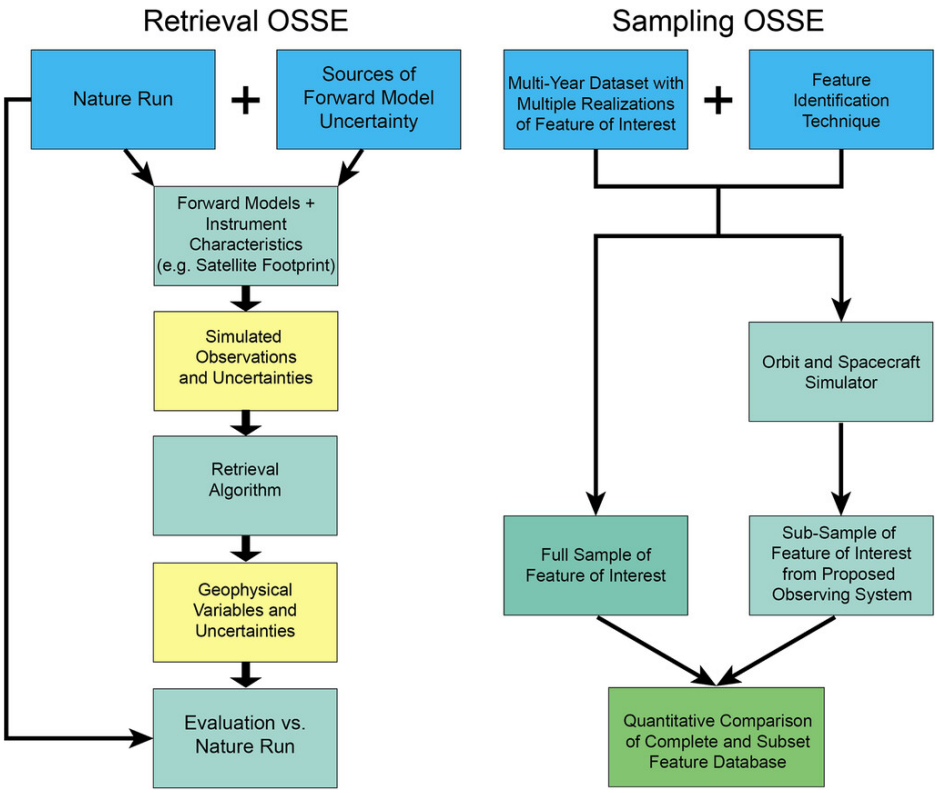


Figure 1.8: Schematic depiction of both sampling and retrieval OSSEs as used in NASA (Zeng et al., 2020).

The traditional forecasting OSSEs make use of Nature Run (also known as synthetic) from all observations are simulated while the sampling OSSEs use an alternate model to simulate only the new observations whose impacts are to be estimated and allow for the examination of actual events in the historical record (Boukabara et al., 2016; Hoffman and Atlas, 2016; Zeng et al., 2020). However, although the historical OSSEs may be useful for understanding specific cases, their use in decision-making for observing system design is cautioned (Zeng et al., 2020).

Modern OSSE frameworks are made up of three main components: (1) a Nature Run (NR), (2) a Data Assimilation System (DAS), and (3) a programme or software to extract "pseudo-observations" from the Nature Run and to include realistic observation errors (Halliwell et al., 2014; Atlas et al., 2015; Zeng et al., 2020). While a NR is a model simulation with high possible resolutions whose output is assumed to be closely representative of the true environmental conditions from a statistical viewpoint, a DAS is a system that combines both observations and model output to produce a more optimal estimate of the evolving state of the observing system (Andersson and Masutani, 2010). For example, OSSEs have been performed to:

- Make design decisions for making a new observing system or network (Mazloff et al., 2018; Kamenkovich et al., 2017).
- Find out whether a new observing system will add value to numerical weather prediction and analysis (Hoffman and Atlas, 2016; Atlas, 1997).

- Investigate the behaviour of DASes and thereby optimally tune these systems in an environment where the ”truth” and hence the system's behaviour is known (Boukabara et al., 2016).

The application of OSSEs in the ocean carbon community has been almost non-existent, particularly in the Southern Ocean, until recently. The first track of published work using the OSSE in the ocean carbon community is by Majkut et al. (2014). Motivated by the limitation of our understanding of the dynamics governing the Southern Ocean CO_2 uptake − mainly due to sparse observations and incomplete model formulation − the authors used OSSEs to show that float-based sampling provides an opportunity for measuring the mean CO_2 fluxes and monitoring the mean uptake over time-scales. Specifically, they used OSSEs to review the model and empirical estimates for CO_2 uptake in the Southern Ocean and then investigate the float-based sampling benefits and strategies.

Recently, a few studies coupled OSSE and empirical mapping approaches to investigate some pressing scientific questions regarding errors in surface ocean pCO_2 estimates and pCO_2 field reconstructions (Gloege et al., 2021; Denvil-Sommer et al., 2021; Valsala et al., 2021). These studies all used ocean model output and assumed that they provide plausible representations of the relationships that exist between pCO_2 and its ancillary variables. Specifically, at a global scale, Gloege et al. (2021) used OSSEs on Earth System Models with SOCAT-ship track data to quantify errors in observation-based estimates of seasonal, decadal, and sub-decadal variability of ocean carbon sink. In their study, the authors also added SOCAT-based samples to assess the impacts on reducing uncertainty. At a regional scale, in the Atlantic Ocean to be specific, Denvil-Sommer et al. (2021) carried out a series of OSSEs with geographical and time locations from three ocean observing platforms including (1) volunteering observing ships, (2) Argo floats, and (3) OceanSITES moorings. Lastly, Valsala et al. (2021) carried out almost similar OSSEs as Denvil-Sommer et al. (2021) in the Indian Ocean to identify potential locations for making surface ocean pCO_2 measurements using moorings and ships.

1.3.2 State of pCO_2 reconstructions in the Southern Ocean

Novel reconstruction methods of surface ocean CO_2 all seek to fill – "reconstruct" – 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^{ocn}) using prognostic ancillary variables. Different numerical and empirical approaches have been used for pCO_2 reconstruction, regionally and globally. Many empirical approaches such as statistical interpolations and regression methods (Iida et al., 2015; Jones et al., 2015; Rödenbeck et al., 2003, 2014) gained attention as alternative methods to ocean biogeochemical models (Lenton et al., 2013) until recently when machine learning approaches have been used increasingly as an alternative (Gregor et al., 2017, 2018; Landschützer et al., 2013, 2014; Landschützer et al., 2015). The ancillary variables by these novel methods are usually satellite-observed and re-analysis-based sea surface temperature,

sea surface salinity, mixed layer depth, chlorophyll-*a*, etc. The feasibility of these extrapolations is based on a known but complex relationship between the surface ocean pCO_2 and the above-mentioned prognostic variables, which is sought through various approximations (Takahashi et al., 1993).

Hindcast simulations with ocean models and observation-based mapping techniques (Hauck et al., 2020; Friedlingstein et al., 2021) are integral to providing a regional picture of the evolution of ocean carbon sink (Gloege et al., 2021). But observations are too sparse in time and space to directly constrain air-sea CO_2 exchange. For example, globally, the 2020 release of the SOCAT database covered only 1.5-2% of all possible monthly $1^\circ \times 1^\circ$ grids in the period 1982-2019 (Gloege et al., 2021). The data sparseness is the most significant in the stormy autumn and winter seasons, especially in remote regions such as the Southern Ocean, requiring substantial extrapolations to map surface ocean pCO_2 and to subsequently fill the sampling gaps (Gregor et al., 2017, 2019; Landschützer et al., 2014). These sampling limitations challenge the ability of mapping methods to accurately reconstruct surface ocean pCO_2. The most popular and current mapping techniques combine unsupervised & supervised machine-learning methods such as the self-organizing map and feed-forward neural network (Landschützer et al., 2016) and mini-batch K-Means & feed-forward neural network or gradient boosting machine(Gregor et al., 2019), to provide continuous monthly estimates of pCO_2. However, existing differences in the results of these pCO_2 reconstruction methods in the Southern Ocean (e.g., see Section 1.1.3) are a real challenge to the direct comparison of pCO_2 estimates to ocean model outputs (Mongwe et al., 2018) as well as the benchmarking of Earth system model-based predictions (Li et al., 2020). This would subsequently influence the assessment accuracy of the regional and global carbon budget as these results are now being used in the global carbon budget (LeQuéré et al., 2016; Friedlingstein et al., 2021).

1.4 Structure of the thesis

The importance of the Southern Ocean climate sensitivity in the regional and global climate was outlined in Section 1.1 which leads to the focus on the importance of improving the accuracy and precision of ocean carbon sink estimates in the region. However, temporal and spatial observational gaps in the Southern Ocean continue to be a problem. This is exacerbated by the lack of a rigorous scale-sensitive sampling strategy for observing platforms in the Southern Ocean to close observational gaps and reduce biases and uncertainties in surface ocean pCO_2 reconstructions.

By scale sensitive sampling strategy, I mean a sampling strategy which addresses limitations including winter and basin-scale sampling gaps as well as seasonal sampling aliases in high-EKE and mesoscale-intensive ocean regions. In increasing CO_2 sampling efforts in the Southern Ocean using autonomous sampling platforms through research programs like SOCCOM and the Southern Ocean Carbon–Climate

Observatory (SOCCO), the community has tried to answer this problem, but the effectiveness of these efforts has not yet been tested. Addressing this aspect of the problem – the sampling scale sensitivity of pCO_2 reconstructions – will allow the community to provide accurate estimates of air-sea CO_2 fluxes in the Southern Ocean. This will contribute to an accurate mean annual global carbon budget which is critical for the trend of the ocean sink feedback on global warming as well as ocean acidification.

1.4.1 Aims

Based on the bulk formulation of air-sea CO_2 fluxes, F_{CO_2} given in Equation 1.1.2, any biases and uncertainties present in surface ocean pCO_2 reconstructions have direct implications in the F_{CO_2} calculation, therefore, hamper the accuracy and precision of the estimation of the Southern Ocean CO_2 sink. In this thesis, I aim to advance understanding of the sampling scale sensitivity of pCO_2 reconstructions and contribute – through a scale-sensitive sampling strategy of observing platforms in the Southern Ocean – to breaking through the proverbial "wall" (see Section 1.1.2) on which the performances of observation-based pCO_2 estimates have been converging.

1.4.2 Outline

Henceforth, this thesis is outlined as follows. Chapter 2 of this thesis presents the data sets and the methodologies common to all the other research chapters. Chapter 3 of the thesis uses a set of semi-idealized OSSEs to discuss whether the scale-sensitive sampling in the Southern Ocean matters in pCO_2 reconstructions in the region. Chapter 4 looks at the sensitivity of two-step pCO_2 reconstructions to clustering choices for the Southern Ocean as a whole. Specifically, in this chapter, I evaluate the clustering methods, the first step in the two-step pCO_2 reconstructions and the contribution that each of the clustering methods or their ensemble make to the bias and root mean square error in the contrasting summer and winter sampling scenarios. Chapter 5 investigates a scale-sensitive integrated pCO_2 observing system for the Southern Ocean in order to "get over the wall" (see Section 1.1.2). Here, I propose a scale-sensitive sampling strategy that addresses the limitations of underway SOCAT ship observations in the whole Southern Ocean and strengthens the existing summer-biased ship-based reconstructions of surface ocean pCO_2 in order to "get over the wall". Chapter 6 reflects on the "big" question of the thesis which is the sampling scale sensitivity of pCO_2 reconstructions in the Southern Ocean. Here, I thus conclude by discussing the limitations of my study and recommendations for future observational and modelling studies.

Chapter 2

Data, modelling frameworks and theory

In this chapter, I present the data sets and the methodologies common and used throughout all the research chapters. Specifically, I start by presenting the technical details of the ocean model data products together with its configurations within the study region (i.e., Southern Ocean) where we explore, process, and visualize the model prognostics variables that are required throughout this thesis. Then, I present the essence of the research questions of this study through the Southern Ocean Seasonal Cycle Experiment (SOSCEx) project. The final sections of the chapter introduces the machine learning (ML) regression frameworks used in this thesis with a specific focus on a class of feed-forward neural network called Multi-layer Perceptron, and on another sophisticated ML regression algorithm from the class of Gradient Boosting Decision Tree called Gradient Boosting Machine. I also explain some fundamental theories and applications these gap-filling methods and describes how to assess the resulting models and optimize them through tuning of their hyperparameters in order to obtain much stable and generalizable models to make predictions. The second last section introduce the general conpception of observing system simulation experiments (OSSEs), while the last section give a thorough decomposition of various sources of potential uncertainties required for the purpose the analyses in this thesis.

2.1 A regional forced ocean model

2.1.1 BIOPERIANT12 model specifications

The ocean model data used in this thesis are from a year-long mesoscale resolving ($\pm$10 km) ocean model simulation. This ocean model is thus a regional configuration, BIOPERIANT12-CNCRUN05A-S (BIOPERIANT12), 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).

Specifically, the BIOPERIANT12 ocean model whose a general configuration is described in Table 2.1, is a high-resolution ($\pm$10 km) forced NEMO-PISCES regional Southern Ocean model. This model simulates daily outputs but only five-daily mean outputs are saved because of the limited computational memory and storage resources in the supercomputing system on which the BIOPERIANT12 model is run and hosted.

Table 2.1: Specifications of BIOPERIANT12-CNCRUN05A-S configuration (the forced NEMO-PISCES regional Southern Ocean model). This table is adapted from Mongwe et al. (2016) and Gregor (2017).

Domain	Regional Circumpolar	30°S to Antarctic continent 77°S	
Horizontal Grid	Resolution $\frac{1}{12^\circ} \times \frac{1}{12^\circ}$	$x = 4322$ $y = 1122$	
Vertical Grid	46 vertical levels $z-$coordinate + partial bottom steps	upper 200 m below 200 m	15 levels 20 levels
Initial Conditions	Started from rest		
	T and S from the World Ocean Atlas		
	Sea ice from ORCA12		
Run Duration	Spin-up Run	1989-1995 1996-2009	
Lateral Boundary	Only northern boundary open	5 day ocean dynamics from ORCA12 1 month tracer climatologies: Alk, DIC, DOC, De, NO_3, O_2, PO_4, Si	
Surface Boundary	Atmospheric Fforcing	ERA Interim	
		10 m u and v	3 hours
		swrad, lwrad	24 hours
		tair, thumi $(t2, q2)$	3 hours
		precio, snow	24 hours
	Damping	S and T are damped to climatologies	
	Runoffs	S and T are included at a frequency of 24 hours River treatments, mixing over upper 10 m	
	Sea surface restoring	Salinity only - using Levitus climatolog	
	penetration of light	RGB formulation with Chlorophyll read from file	
Bottom Boundary	Non-linewar bottom friction		
Model Numerics	Ocean tracer advection scheme: total variation friction (TVD)		
	PISCES tracer advection scheme: Monotonic Upstream-centered Scheme for Conservation Laws (MUSCL)		
	Eddy-resolving		
	Laplacian lateral diffusion for tracers along isoneutral surfaces		
	Biplacian lateral diffusion of momentum along geopotential surfaces		
	Vertical diffusion is handled by TKE scheme, with the source below the ML		
	Langmuir parameterisation, surface mixing length scale is a function of wind		

Due to the computational memory and storage resource limitations of the supercomputing system that I used for the study, the high-resolution ($\pm$10km) ocean model run with daily output was only available for one year, 2009, the last year run of the BIOPERIANT12 ocean model (see Table 2.1). However, a specific model year is not here necessary because the thesis takes a semi-idealized approach that assumes a single year to be the benchmark to explore the sensitivity of pCO_2 reconstructions to the temporal and spatial observation-scale gaps stemming from observing platforms such as ships and autonomous vehicles.

2.1.2 Prognostic variables of interest and processing

The BIOPERIANT12 ocean model has many prognostic variables including two phytoplankton types (diatoms and nanophytoplankton) and a description of the marine carbonate in the model. However, we focus only on the variables that are of particular interest for our study. These variables include the partial pressure of carbon dioxide (pCO_2) at the surface ocean (pCO_2^{ocean}), and its well-known physical and biogeochemical drivers (Takahashi et al., 1993) including the sea surface temperature (SST), sea surface salinity (SSS), mixed layer depth (MLD), chlorophyll-a (Chl-a) concentration. Their characterization is presented with more details in Table 2.2. The pCO_2 reconstruction methods try to establish a non-linear relationship between the surface ocean pCO_2 (or its variants such as the air-sea CO_2 gradient ΔpCO_2), and a set of proxy variables that may vary from one method to another (Landschützer et al., 2013; Rödenbeck et al., 2015; Gregor et al., 2017, 2019). Moreover, the common assumption is these drivers capture the major modes of spatio-temporal variability of pCO_2^{ocn}.

Table 2.2: Characteristics of the BIOPERIANT12 prognostic variables of interest and basic data processing steps used for the proxy variables of the surface ocean pCO_2. It is important to not that the the NEMO-PISCES sign of the gradient, ΔpCO_2, is reversed in the thesis; that is, positive(+) is a flux from the atmosphere (atm) to the ocean (ocn) unlike what is usually used in the literature (cf. Equation 1.1.1). Abbreviations are used throughout the thesis.

Variables	Abbreviations	Processing	Date range	Resolutions	
				spatial	temporal
Air-sea pCO_2 gradient	ΔpCO_2	Model simulations $pCO_2^{atm} - pCO_2^{ocn}$	1 year	$\left(\frac{1}{12}\right)^\circ$	daily
Atmospheric (atm) pCO_2	pCO_2^{atm}	in-situ			
Surface ocean pCO_2	pCO_2 or pCO_2^{ocn}	$pCO_2^{atm} - \Delta pCO_2$			
Sea surface temperature	SST	Model simulations			
Sea surface salinity	SSS	Model simulations			
Mixed layer depth	MLD	Model simulations $\log_{10}$ transformation			
Nano chlorophyll	NChl	Model simulations			
Datom chlorophyll	DChl	Model simulations			
Chlorophyll-*a*	Chl-*a*	DChl + NChl $\log_{10}$ transformation			
Day of the year (DoY)	$\text{DoY}_{\cos}$	$\cos\left(\text{DoY} \times \frac{2\pi}{365}\right)$			
	$\text{DoY}_{\sin}$	$\sin\left(\text{DoY} \times \frac{2\pi}{365}\right)$			

The BIOPERIANT12 ocean model run outputs ΔpCO_2 before the surface ocean pCO_2 data can be derived through preprocessing as summarized in Table 2.2. Similarly, the total chlorophyll-*a* concentration was derived from nano and diatom phytoplankton chlorophyll concentation (diatom + nano phytoplankton).

For empirical estimation of pCO_2, outputs of some prognostic variables of interest such as MLD and Chl-*a* had to be preprocessed. They underwent a $\log_{10}$ transformation in order to return a distribution closer to a normal distribution (Maritorena et al., 2010; Holte et al., 2017). In practice, existing reconstruction methods have been using MLD climatology as a proxy variable (Gregor et al., 2019; Gloege et al., 2021). This enables a smoothing of the data and thus reduces the uncertainty from MLD information. However, in the OSSE setting, using MLD from the model rather than a climatology is likely an advantage compared to the existing methods that use MLD climatology. The advantage of including proxy variables such as MLD and Chl-*a* is that the model is providing constraints which might not be available from real-world observations. The temporal coordinate of the data is also included as proxy through a variable transformation that aims to preserve the seasonality of the data. This seasonality preservation was done by transforming the day-of-year say j as in Gregor et al. (2017) (see Table 2.2). The maps of annual means of the surface ocean pCO_2 and its key proxies used in the modelling are shown in Figure 2.1. Spatial coordinates (latitude, longitude) are not included in the pCO_2 predictors like in (Gregor et al., 2017, 2018) and many other studies used in (Rödenbeck et al., 2015). This is because according to (Gregor et al., 2017), coordinate variables do not drive mechanistic changes in pCO_2.

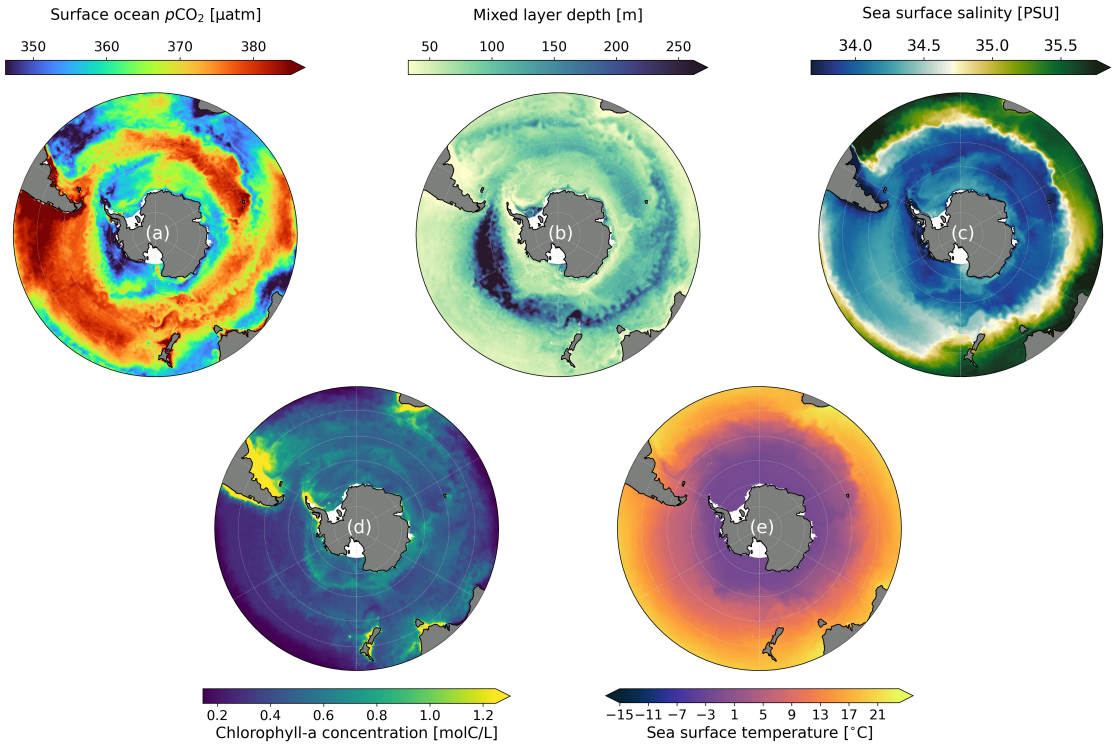


Figure 2.1: Southern Ocean annual mean of the BIOPERIANT12 (a) surface ocean pCO_2, (b) mixed layer depth (MLD), sea surface salinity (SSS), (d) total chlorophyll-a concentration (diatom + nano phytoplankton), and (e) sea surface temperature (SST).

2.2 SOSCEx project

According to many studies, the seasonal cycle is known as the strongest mode of natural CO_2 variability in the Southern Ocean and also the one that most strongly links climate and ocean ecosystems (Mongwe et al., 2018; Gruber et al., 2009, 2019). Its characteristics are largely shaped by high-frequency intra-seasonal modes defining the response modes in physics and biogeochemistry components (Mongwe et al., 2016, 2018). Therefore, the Southern Ocean Seasonal Cycle Experiment (SOSCEx) project - an initiative of the Southern Ocean Carbon and Climate Observatory (SOCCO), a research programme led by the Council of Scientific and industrial Research (CSIR) - was launched in 2013. The SOSCEx goal was to explore the nature and links in dynamics and scale sensitivities of atmospheric forcing, CO_2 fluxes, and primary production, with a particular focus on the seasonal cycle mode as a test for the climate sensitivity of earth systems models in respect of the evolution of both atmospheric CO_2 and ocean ecosystems in the 21st century (Swart et al., 2012; Monteiro et al., 2010, 2015).

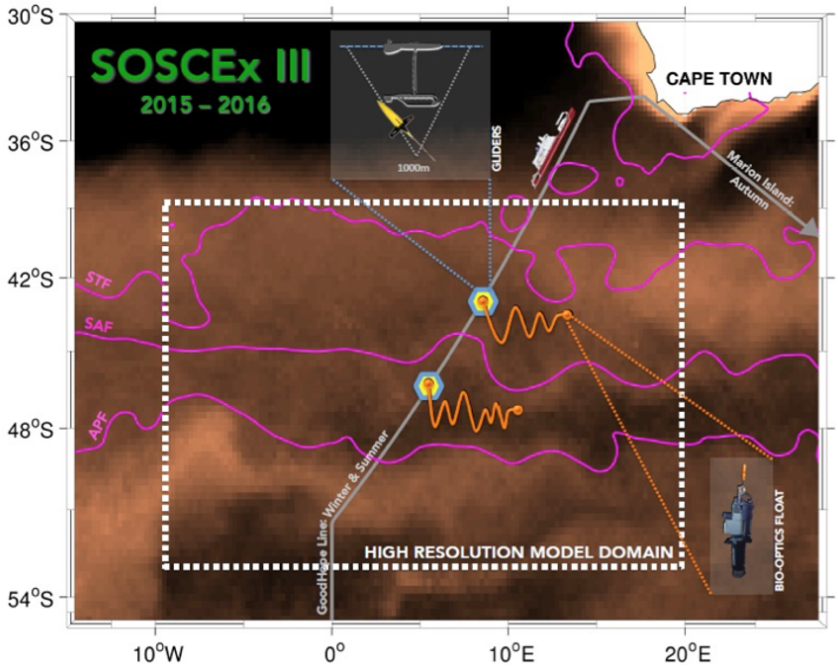


Figure 2.2: Schematic view of the observing strategy the third phase of SOSCEx project. The figure illustratesthe use of multiple ocean sampling tools including ships, gliders, floats, and numerical models. The hexagonal patterns (blue-yellow) depict the twined glider deployments; the orange curve shows the Lagrangian float sampling trajectories, while the high-resolution modellingdomain is represented with the white dashed line. In magenta lines are the average locations of the oceanic fronts shown as derived from satellite altimetry data, whereas the underlying shading depicts the mean summer chlorophyll-a concentration in the region with lighter shading=high Chl-a areas. (Source: https://socco.org.za/news/plans-underway-for-soscex-iii/).

As illustrated in Figure 2.2, the novel aspect of the third phase of that project (SOSCEx III, 2015-2018) was the integration of a multi-platform approach. This 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 ocean biogeochemical models, hence the high-

resolution forced NEMO-PISCES regional Southern Ocean model, BIOPERIANT12.

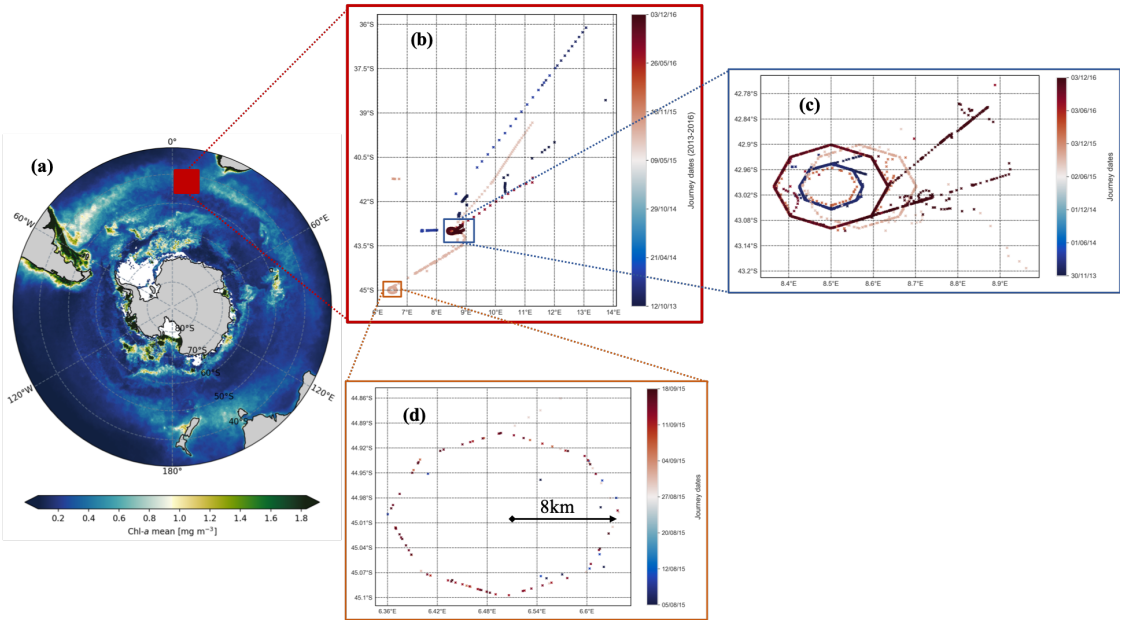


Figure 2.3: This figure depicts our study area and the main characteristics of the site as follows. **(a)** This is the study region with the sampling location of the gliders (red box) overlaid on the regional mean SeaWIFS chlorophyll-*a* (Chl-*a*). **(b)** Entire sampling domain of the Liquid Robotics Carbon Waveglider, dubbed WG SV2, from deployment in the North-west quadrant to its long term pseudo-mooring polygon sampling pattern in the SAZ, located at about 1300 km South-west of Cape Town, South Africa. The blue **(c)** and orange **(d)** boxes represent the zoom of the polygon sampling pattern of the Wave Glider (WG) in site 1 and site 2, respectively, with the radius of the polygon indicated ($\sim$ 8 km). The centre of the polygon represents the mean location of the Seaglider profiling to 1000 m depth. The color shading in **(b)**, **(c)** and **(d)** depicts the temporal evolution of the experiment, that we called here WG journey dates.

2.3 Machine learning frameworks and theory

In a data analytic viewpoint, the problems we are addressing in this thesis are tackled as a regression predictive modelling problem. Simplistically, this means a modelling problem that involves interpolating a real-valued quantity (such as the surface ocean pCO_2) by leveraging the prediction skill of a Machine Learning (ML) algorithm trained on the data associated with the problem. When it comes to solving a ML problem, often the hardest task can be finding suitable algorithms for the job. The main reason is that different methods are better suited for different types of data and different problems as described by the flowchart in Figure 2.4.

2.3.1 Learning a ML model

It is important to keep in mind that the algorithm does the learning from training data, and then the model contains learned funtional relationships as depicted on Figure 2.4. The model itself may be a handful of numbers, and ways of using those numbers to relate input (features/proxies) to output (target). In predictive modelling, there are three concepts that can be considered as fundamental to answering the underlying ques-

tion. The first concept is the Sample Data that represents the collected data or information that describes the

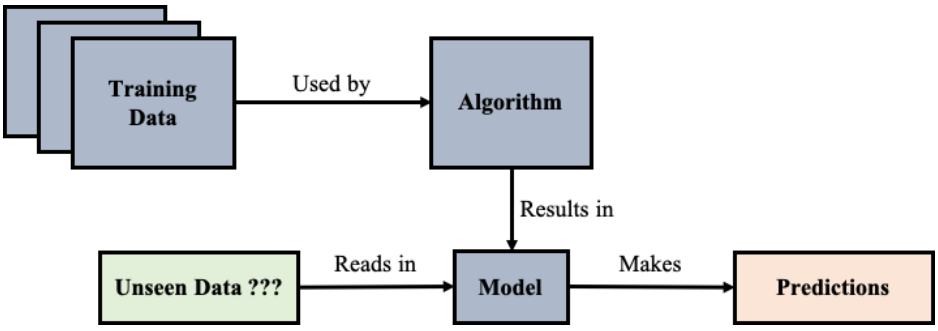


Figure 2.4: This figure depicts the creation of a predictive model from training data and an algorithm. This resulting model can then be used to make predictions.

problem with the known relationship between the inputs and the outputs; for example, variables such as sea surface temperature (SST), sea surface salinity (SSS), surface chlorophyll-a (Chl-a) and mixed layer depth (MLD) serve as proxies or drivers for known processes that modulates the surface ocean pCO_2 (Sarmiento and Gruber, 2006). The second concept is to Learn a Model that represents the algorithm used on the sample data to create a model that can later be used over and over again. The third and final concept we consider here is Making Predictions, that is, the use of the previously learned model on new input data for which the outputs are unknown. Previous studies such as Landschützer et al. (2013), Gregor et al. (2019), and Denvil-Sommer et al. (2019) leveraged similar concepts in order to reconstruct the global ocean maps of the surface ocean pCO2 and subsequently that of air-sea CO_2 fluxes.

Function estimation

In general, in machine learning applications, particularly in the classical supervised leaning setting like that of this thesis, the problem resulting in the problem of function estimation. The learning being supervised restricts us, researchers and practitioner, on avaibility of data, especially the target data, which can be very costly to obtain (e.g.; surface ocean pCO_2). Let us consider the dataset $(x,y)_{i=1}^{N}$, where N is the size of the dataset, $x = (x_1, \cdots, x_p)$ referes to the ancillary input variables and y to the corresponding values of the response variable such as the surface ocean pCO_2. The goal is thus to reconstruct the unknown functional relationship $f : x \rightarrow y$ so that the resulting estimate $\hat{f}(x)$ minimizes some specified loss function $\Phi(y, f)$ (Natekin and Knoll, 2013); that is,

$$y = \hat{f}(x) = \underset{f(x)}{\operatorname{argmin}}\, \Phi(y, f(x)) \tag{2.3.1}$$

Not knowing neither the true functional dependence of $f(x)$, nor the exact form the resulting estimate $\hat{f}(x)$, the estimation problem can be written in terms of expections like in mathematical statistics (Hastie et al., 2009; Natekin and Knoll, 2013). Therefore, the equivalent formulation of Equation 2.3.1 will be to minimize $\mathbb{E}_y[\Phi(y, f(x))]$ – the expected loss over the response variable (y) – conditioned on the explanatory data x

as follows:

$$\hat{f}(x) = \underset{f(x)}{\operatorname{argmin}} \, \mathbb{E}_x \left\{ \mathbb{E}_y \left[\Phi(y, f(x)) \right] | x \right\}. \tag{2.3.2}$$

In many applications, this function estimation problem is tractable, which means the function search space is restricted to a parametric family of functions $f(x, \theta)$. Thus, the function optimization problem (Equation 2.3.2) would become a parameter estimation problem (cf. Equation 2.3.3).

$$\hat{f}(x) = \underset{\theta \in \Theta}{\operatorname{argmin}} \, \mathbb{E}_x \left\{ \mathbb{E}_y \left[\Phi(y, f(x, \theta)) \right] | x \right\}, \tag{2.3.3}$$

where Θ is the parameter space, and

$$\hat{f}(x) = f\left(x, \hat{\theta}\right). \tag{2.3.4}$$

Now, the resulting function estimate $\hat{f}$ consititutes the model and can be used to make predictions $\hat{f}(\tilde{x})$ on a new data $\tilde{x}$ of the explanatory variable x. Sections 2.3.2 and 2.3.5 give an overview on how the optimization in this function estimation work in practice for the two machine learning regression algorithms used in this thesis.

2.3.2 Feed-forward neural networks

Model description and basic formulation

In this section, we present some key theoretical concepts and applications of the class of neural netwok algorithms that are used in this thesis. More precisely, we used a class of feed-forward artificial neural network called Multi-layer Perceptron (MLP) that consists of at least three layers of nodes/units: an input layer, a hidden layer and an output layer. Each unit is a neuron that uses a non-linear activation function. In artificial neural network, the activation function of a unit defines the output of that unit given an input or a set of inputs.

As many other Machine Learning methods, in the training phase, MLP makes use of a widely used supervised learning techniques called back-propagation. Based on the graph above (see Figure 2.5), let's consider for example that fully-connected network with one hidden layer and no biases, trained to predict an output value y (e.g. pCO_2^{ocn}, the surface ocean pCO_2) from the feature variable or proxy x (e.g. salinity, sea surface temperature, etc.) using the Euclidean metric – defining the distance between a ground-truth data point and the estimated one – for the loss function.

Theoretical analysis

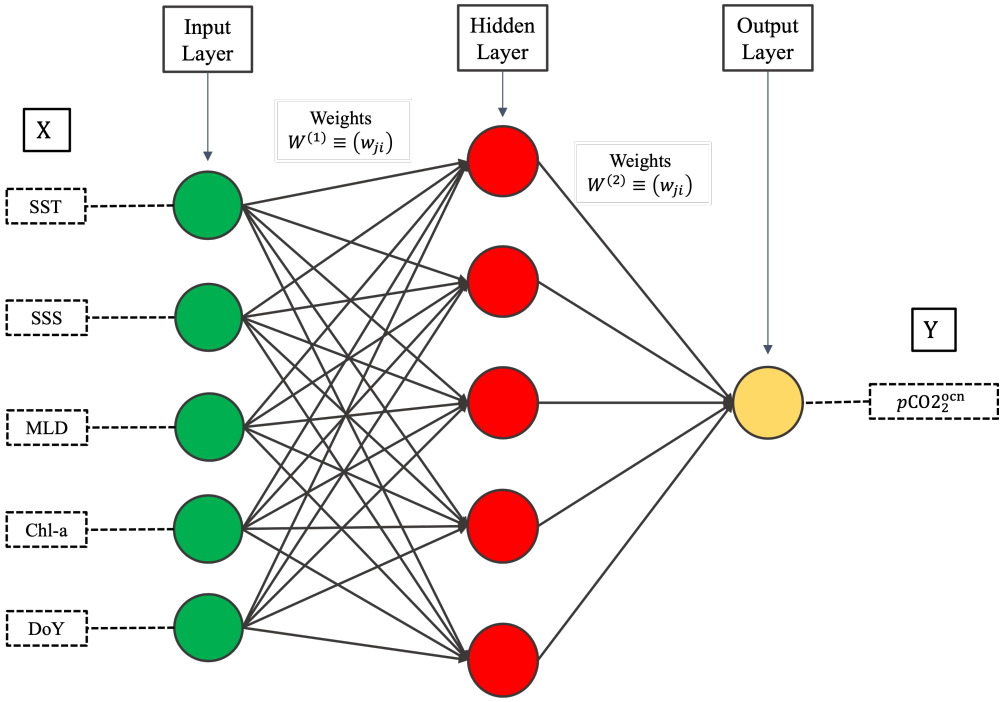


Figure 2.5: Depiction of a typical example of the architecture or graph of a simple Multi-layer Perceptron network, a 2-layer neural network as the input layer does not usually count when counting the number of layers in a network. Its multiple layers and non-linear activation function distinguish MLP from linear perceptron.

Using the non-linear activation function called Rectified Linear Unit (ReLU) and defined as follows:

$$\text{ReLU: } x \longmapsto \max(0,x) = \begin{cases} x \text{ if } x \geq 0 \\ 0 \text{ else} \end{cases},$$

our MLP example can be set up as follows:

$$\mathbf{Y}^{(i)} = \max\left(0, \mathbf{X}^{(i)} w^{(1)}\right) w^{(2)},$$

where $i = 1,2,\cdots,n$ with n the total number of samples and corresponds to the i-th example, and: $\mathbf{Y^{(i)}} = \begin{bmatrix} y_1^{(i)} \\ y_2^{(i)} \\ \vdots \\ y_c^{(i)} \end{bmatrix}$ is the i-th output target with c output values (for our application $c = 1$), $\mathbf{X^{(i)}} = \begin{bmatrix} x_1^{(i)} \\ x_2^{(i)} \\ \vdots \\ x_p^{(i)} \end{bmatrix}$ is the i-th

input data with p features (for our application $p = 4$), and $w^{(\ell)} = \begin{bmatrix} w_1^{(\ell)} \\ w_2^{(\ell)} \\ \vdots \\ w_{p_\ell}^{(\ell)} \end{bmatrix}^T$ is the weight vector of the ℓ-th layer with $\ell = 1, 2$, and p_ℓ is the number of weights . Subsequently, we are going to consider only one sample (x, y) where $c = 1$ as in our application the output has only one target, the surface ocean pCO_2 we denoted by pCO_2^{ocn}.

2.3.3 Definition. The back-propagation algorithm in a graph consists of two important phases described as follows.

1. **Forward pass:** in this first phase, for each node, we compute local partial derivatives of the loss function given inputs.
2. **Backward pass:** in this phase, we apply chain rule from the end to each parameter, update these parameters with gradient descent using the current upstream gradient and the current local gradient, and then compute upstream gradient for the backward nodes.

During the **Forward Pass**, given the inputs x (such as sea surface temperature, sea surface salinity, mixed layer depth and chloryphyll), the forward pass with local partial derivatives of output (loss) is structured as illustrated on the Flow Diagram (Table 2.3).

Table 2.3: Flow diagram depicting the forward pass with local partial derivative of the loss function.

$$x \rightarrow \boxed{\overset{w^{(1)}}{\underset{\downarrow}{z^{(1)} = x^T w^{(1)}}}} \rightarrow \boxed{h^{(1)} = max\left(0, z^{(1)}\right)} \rightarrow \boxed{\overset{w^{(2)}}{\underset{\downarrow}{z^{(2)} = h^{(1)T} w^{(2)}}}} \rightarrow \boxed{L\left(z^{(2)}, y\right) = \left(z^{(2)} - y\right)^2}$$

$$\frac{\partial z^{(1)}}{\partial w^{(1)}} = x \qquad \frac{\partial h^{(1)}}{\partial z^{(1)}} = \begin{cases} 1 \text{ if } z^{(1)} > 0 \\ 0 \text{ else} \end{cases} \qquad \frac{\partial z^{(2)}}{\partial w^{(2)}} = h^{(1)} \qquad \frac{\partial L}{\partial z^{(2)}} = 2\left(z^{(2)} - y\right)$$

$$\frac{\partial z^{(1)}}{\partial x} = w^{(1)} \qquad \frac{\partial z^{(2)}}{\partial h^{(1)}} = w^{(2)}$$

Afterwards, the **Backward Pass** consists of computing the gradient of the loss, $L\left(z^{(2)}, y\right)$, given each parameters vectors by applying the chain rule from the loss downstream to the parameters:

- **Chain Rule on** $w^{(1)}$:

$$\frac{\partial L}{\partial w^{(1)}} = \frac{\partial L}{\partial z^{(2)}} \frac{\partial z^{(2)}}{\partial h^{(1)}} \frac{\partial h^{(1)}}{\partial z^{(1)}} \frac{\partial z^{(1)}}{\partial w^{(1)}} = 2\left(z^{(2)} - y\right) w^{(2)} \left(\begin{cases} 1 \text{ if } z^{(1)} > 0 \\ 0 \text{ else} \end{cases} \right) x$$

- **Chain Rule on $w^{(2)}$:**

$$\frac{\partial L}{\partial w^{(2)}} = \frac{\partial L}{\partial z^{(2)}} \frac{\partial z^{(2)}}{\partial w^{(2)}} = 2\left(z^{(2)} - y\right) h^{(1)}$$

After applying the chain rule from the loss downstream to the parameters, we derived the final expression using the Flow Diagram of the forward pass (Table 2.3).

2.3.4 Remark. The forward pass above operates over vectors, matrices and/or tensors. We are thus dealing with matrix multiplication. This means in the computations of the derivatives for backwark pass, we are not dealing with regular derivative, but with vector/matrix derivatives.

2.3.5 Gradient boosting machines

Gradient Boosting Machines (GBMs) is a families of machine learning (ML) techniques that are built of decision trees and have demonstrated considerable success in a wide range of practical ML applications (Natekin and Knoll, 2013).

Algorithm description

The Light Gradient Boosting Machine (LightGBM) technique, developed by Ke and colleagues in 2017, is a widely-used ML algorithm, due to its efficiency, accuracy, and interpretability. It is a ML framework based on decision trees and the idea of gradient boosting, and developed for predictive regression and classification problems or data mining tasks such as ordering (Ke et al., 2017). The concept gradient boosting can be seen as an optimization algorithm on a suitable cost function (Breiman, 1998). LightGBM framework, as a relatively new technique, is introduced in detail here. It produces a prediction model in the form of an ensemble of weak prediction models, typically decision tree learners that increases the efficiency of the model and reduces memory usage during training. The decision tree learning algorithm goes from observations about an item to conclusions about the item's target value. LightGBM builds the model in a stage-wise fashion like other boosting methods do, and it generalizes them by allowing optimization of an arbitrary differentiable loss function.

Theoretical analysis

Primarily, all GBDT (Gradient Boosting Decision Tree) techniques such as eXtreme Gradient Boosting (XGB) make use of histogram-based algorithm to produce a prediction model in the form of an ensemble, but these algorithms presents some limitations due to the efficiency and scalability which are still unsatisfactory when the feature dimension is high and data size is large (Ke et al., 2017). Thus, in order to fulfill these limitations of histogram-based algorithm, LightGBM uses two novel techniques: Gradient-based One Side Sampling (GOSS) and Exclusive Feature Bundling (EFB). These two techniques (GOSS and EFB) form

the key characteristics of the LightGBM algorithm., and both combined together makes the resulting model usually works better compared to other GBDT frameworks such as XGB, Random Forest and Extra Trees (Ke et al., 2017). However, all these algorithms are variants of the original Gradient Boost algorithm developed by (Friedman, 2001) and summarized below.

Algorithm 1: Friedman's Gradient Boost algorithm

Input:

- $N \geq 1$ // Size of the data
- M // Number of iteration
- $\Phi(y, f)$ // Choice of the loss-function
- $h(x, \theta)$ // Choice of the base-learner model

Data: $(x, y)_{i=1}^{N}$ // Training data: $x = [\mathtt{SST}, \mathtt{SSS}, \mathtt{Chl\text{-}a}, \cdots]$; $y = \left[pCO_2^{ocn}\right]$

Algorithm:

1 $\hat{f}_0 = \text{const}$ // Initialization of $\hat{f}_0$ with a constant

2 **for** $t = 1$ *to* M **do**

3 Compute the negative gradient $-g_t(x)$

4 Train a new base-learner function $h(x, \theta_t)$

5 Find the best gradient descent step-size γ_t:

$$\gamma_t = \underset{\gamma}{\operatorname{argmin}} \left\{ \sum_{i=1}^{N} \Phi\left(y_i, \hat{f}_{t-1}(x_i) + \gamma h(x_i, \theta_t)\right) \right\}$$

6 Update the function estimate:

$$\hat{f}_t = \hat{f}_{t-1} + \gamma_t h(x, \theta_t)$$

Output: $\hat{f} = \hat{f}^M = \sum_{t=1}^{M} \hat{f}_{t-1}$ // Output ensemble learner

More specifically, LightGBM has the following advantages compared to the standard GBDT learning frameworks: faster training speed and higher efficiency, lower memory usage, better accuracy, support of parallel and GPU learning, and have capability of handling large-scale data. Benefiting from these advantages, LightGBM achieves state-of-the-art performances in many ML tasks, and is used in many winning solutions of ML Grand Challenges as we can see the list from https://github.com/microsoft/LightGBM/tree/master/examples.

2.3.6 ML regression metrics

Although the choice of performance measure may seem straightforward and objective, it is often difficult to choose a performance measure that corresponds well to the desired behavior of the system; that is, usually

this performance measure is specific to the task being carried out by the system (Goodfellow et al., 2016). We define the performance metric as any approach used to quantitatively assess or evaluate the performance of a model by measuring how good the model is in terms of estimating the surface ocean pCO_2. To assess the skill of pCO_2 reconstructions we use a series of four statistical metrics including 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).

Mean bias error (MBE)

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:

$$\boxed{\text{MBE} = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)}, \tag{2.3.5}$$

where n is the data size or number of samples, $\hat{y}_i$ is the model prediction of the i-th target varible y_i (i..e., i-th value of pCO_2^{ocean}), for $i = 1, 2, \cdots, n$.

Mean absolute error (MAE)

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:

$$\boxed{\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|\hat{y}_i - y_i|}. \tag{2.3.6}$$

Root mean squared error (RMSE)

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:

$$\boxed{\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2} = \sqrt{\text{MSE}}}, \tag{2.3.7}$$

where MSE is simply the mean square error. For squaring individual errors, that is $e_i = \hat{y}_i - y_i$ for $(i = 1, 2, \cdots, n)$, the stated rationale is usually to "disconnect the sign" of e_i so that the magnitudes of errors

influence the average error, MSE.

Pearson's correlation coefficient (r)

In order to quantify the strength of the linear association between the pCO_2^{ocean} estimates (i.e., $\hat{y}_i$ for $i = 1, 2, \cdots, n$) and observations/known truth (i.e., y_i for $i = 1, 2, \cdots, n$), the Pearson's correlation cofficient (r) is used. Its computing is formulated as follows:

$$\boxed{r = \frac{1}{(n-1)\,\sigma_y \sigma_{\hat{y}}} \sum_{i=1}^{n} (y_i - \bar{y}_i)\left(\hat{y}_i - \bar{\hat{y}}_i\right)}, \tag{2.3.8}$$

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 observations 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.

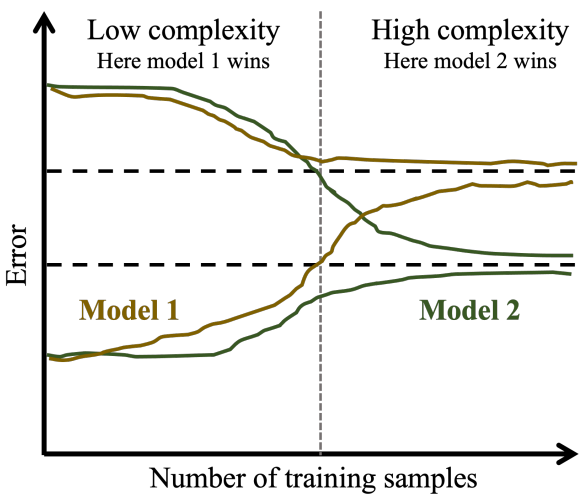
2.3.7 ML model assessment

In machine learning applications, the underlying goal is to achieve models that generalize; that is, perform well on never-before-seen data. However, we can only control well that which we can observe. This brings in the tension between optimization and generalization. By optimization, we refer to the process of adjusting a model to get the best performance possible on the training data (hence the learning in machine learning), whereas generalization refers to how well the trained model performs on data it has never seen before. Thus, the goal of this game is to get good generalization, of course, but we don't control the generalization as we can only adjust the model based on its training data.

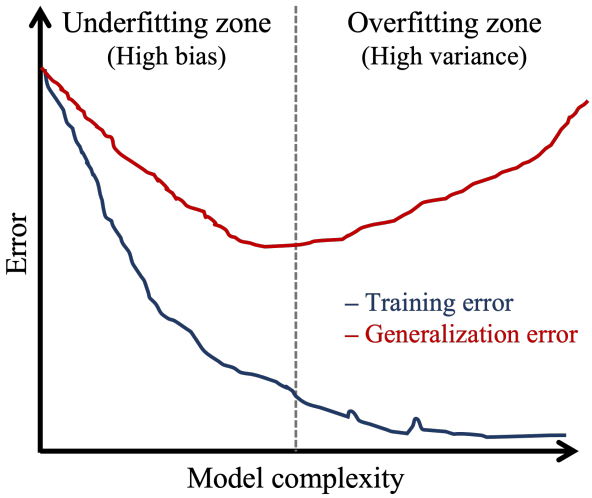
Underfitting vs overfitting

When not handled correctly, this situation can lead to either underfitting or overfitting: The model underfits when we have poor performance on the training data and poor generalization on unseen data, whereas it overfits when we have good performance on the training data, and poor generalization on the unseen data Hastie et al., 2009.

Practically, the training error (in-sample error: error calculated on the training set) actually increases, while the generalization error, that is, the error on never-before-seen data (out-of-sample error), decreases (Figure 2.6a). The two errors converge to an asymptote. Models with different complexity have asymptotes at different error values. In Figure (2.6a), Model 1 converges to a higher error rate than Model 2 because Model 1 is not complex enough to capture the structure of the dataset. However, if the amount of training



(a) Error against training sample size.



(b) Error against model complexy.

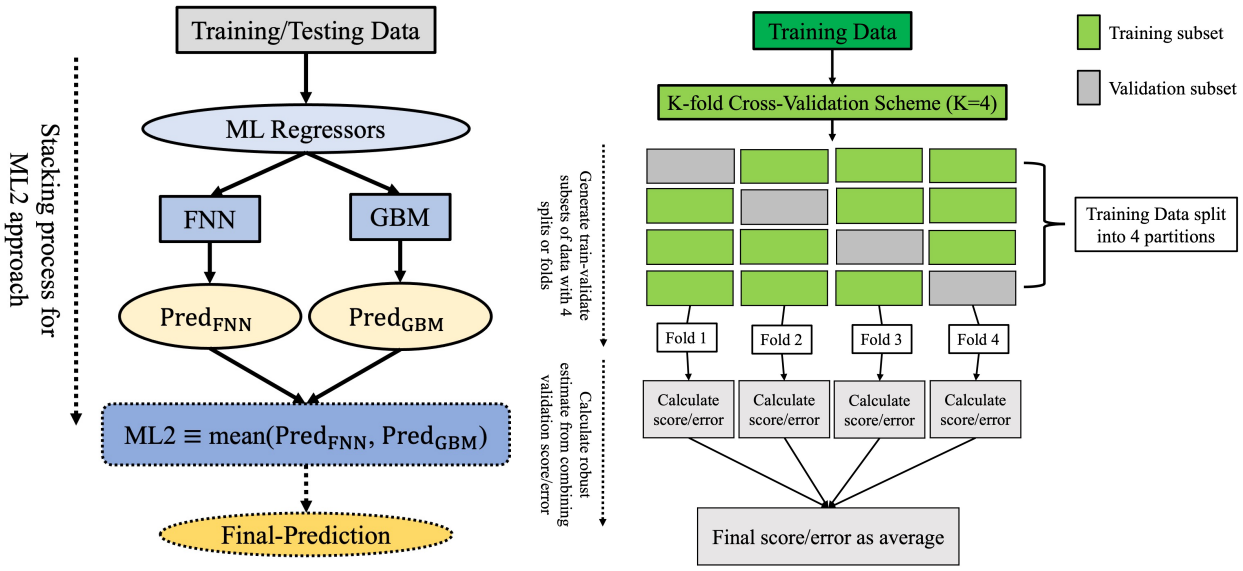
Figure 2.6: A schematic illustration of both underfitting and overfitting scenarios for two different machine learning models (Model 1 and Model 2) used for the same problem. Note that both models can stem from the same machine learning algorithm (cf. Section 2.3.1). **(a)** shows what happens as the number of training samples is increased with a fixed model; **(b)** shows the change in training and generalization errors as the model complexity is increased, and the size of the training set is held constant.

data available is less than a certain threshold, then the less complex Model 1 wins. This regime is important because often the amount of training data is fixed and it is just not possible to obtain any more training data. In the second scenario (Figure 2.6b) the training error decreases monotonically with the model complexity, but the generalization error first falls and then increases. This occurs because by choosing a progressively complex model, at some point, the model begins to overfit the training data; that is, given the larger number of degrees of freedom in a more complex model, the model begins to adapt to the noise present in the dataset, which hurts generalization error.

Addressing overfitting: cross-validation

Overfitting happens in every machine learning (ML) problem and knowing how to deal with it is essential to validate a machine learning model (Hastie et al., 2009; Chollet et al., 2018). Ideally, if we had enough data, we would set aside a validation set and use it to assess the performance of the prediction model (Hastie et al., 2009). Since data are often scarce, this might result in validation scores having a high variance with regard to the validation split, which would prevent us from reliably evaluationg the model. To finesse the challenge, the K-fold Cross-validation (K-fold CV) approach (Hastie et al., 2009) is used (probably the simplest and best common practice).

The K-fold CV approach is illustrated in Figure (2.7). More specifically, its involves randomly dividing the set of training data points into K groups, or folds, of approximately equal size. The first fold is treated as a validation set, and the model is trained on the remaining K-1 folds. The training score, usually the root



(a) Ensemble average learning. **(b)** K-fold cross-validation: K-fold CV.

Figure 2.7: Schematic flow diagram of a two-member ensemble method (ML2) and the K-fold CV procedure. **(a)** shows the schematic representation of the stacking process of a two machine learning algorithms (e.g., FNN and GBM,) that make up the ML2 method; **(b)** shows the schematic flow diagram of the K-fold CV illustrated with K = 4. A set of n observations (training data size) is randomly split into 4 non-overlapping groups/folds. Each of these fourths acts as a validation subset (shown in gray), and the remainder as training subsets (shown light green). The final test score/error is estimated by averaging the 4 resulting mean square error (MSE) estimates.

mean squared error (RMSE) or simply the mean square error (MSE), is then computed on the data points in the held-out fold. This procedure is repeated K times; each time, a different group of data points is treated as a validation set. This process results in K estimates of the test error say, $\text{RMSE}_1, \text{RMSE}_2, \cdots, \text{RMSE}_\text{K}$. The K-fold CV score estimate is computed by averaging these values (Hastie et al., 2009; Gareth James et al., 2014) as follows:

$$\boxed{\text{CV}_\text{K} = \frac{1}{\text{K}} \sum_{k=1}^{\text{K}} \text{RMSE}_k}. \tag{2.3.9}$$

Hyper-parameter optimization (HOP)

Usually, ML models have important parameters which cannot be directly estimated from the data. In general, these parameters are called hyper-parameters, and their optimization is crucial to finding the final model. For example, the *hidden layer size* and the *learning rate* are popular in a FNN model, and the *bagging fraction* and the *depth of tree* in a GBM model (see Table A.1 for more details). Thus, given a set of ML models $\{f(x, \theta)\}$ indexed by a vector β of hyper-parameters, with $x \in \mathbb{R}^p$ (where $\mathbb{R}$ is the set of real numbers and p is the number of features), denote by $\hat{f}(x, \theta)$ the θ-th model trained with the k-th fold and $\text{RMSE}_k\left(\hat{f}, \theta\right)$ the test error (see Equation 2.3.9) associated with $\hat{f}(x, \theta)$. Then for this set of models we defined

$$\boxed{\text{CV}_\text{K}\left(\hat{f}, \theta\right) = \frac{1}{\text{K}} \sum_{k=1}^{\text{K}} \text{RMSE}_k\left(\hat{f}, \theta\right)}. \tag{2.3.10}$$

The function $\mathrm{CV_K}\left(\hat{f},\theta\right)$ provides an estimate of the test error curve. Therefore, the HOP results in minimizing this function (Equation 2.3.10), which also results in finding the tuning parameter $\hat{\theta}_\mathrm{K}$ that satisfies Equation (2.3.11).

$$\boxed{\hat{\theta}_\mathrm{K} = \underset{\theta\in\Theta}{\operatorname{argmin}}\left\{\mathrm{CV_K}\left(\hat{f},\theta\right)\right\}}, \tag{2.3.11}$$

where Θ is the set of the model (f) hyper-parameters. Therefore, the final chosen model is $f\left(x,\hat{\theta}_\mathrm{K}\right)$, which is then trained on all the training dataset, and useed for making predictions on out-of-sample data.

2.4 Regression modelling of surface ocean pCO_2

The dependence of CO_2 partial pressure at the surface ocean (pCO_2) on its main proxies (Takahashi et al., 1993) including the sea surface temperature (SST), sea surface salinity (SSS), mixed layer depth (MLD), chlorophyll-*a* (Chl-*a*), and time as day of the year (DoY, see Table 2.2) is expressed as a non-linear function:

$$p\mathrm{CO2}_2^{ocn} = function\left(\mathrm{SST},\mathrm{SSS},\log_{10}\left(\mathrm{MLD}\right),\log_{10}\left(\text{Chl-}a\right),\mathrm{DoY}_{\sin},\mathrm{DoY}_{\cos}\right). \tag{2.4.1}$$

Spatial coordinates including latitude and longitude data are excluded from this equation in the belief that the combination of SST, SSS, MLD and Chl-*a* provides sufficient spatial information (Zeng et al., 2017). Additionally, having circular properties, these coordinates cannot be used directly. For instance, the longitude value -180° is geographically connected to the longitude value 180°, but they appear to be extreme longitude values to numerical models. Other studies, especially in the Surface Ocean CO_2 Mapping (SOCOM) community (Rödenbeck et al., 2015) tried to circumvent this problem through cosine- and sine-transformed components (Zeng et al., 2015; Gregor et al., 2017, 2018) they could unintentionally enhance the influence of spatial coordinates on pCO_2 estimates. According to Gregor et al. (2017), ML algorithms applied globally struggle to reconstruct surface ocean pCO_2 with reduced biases and uncertainties unless spatial coordinates are included as feature variables. This poor reconstruction may stem from the fact that pCO_2 may respond inconsistently to observable feature variables (such as SSS, SSS, MLD, and Chl-*a*) in different regions as it is impossible to observe all feature drivers of pCO_2.

In this thesis, only the time coordinate data is retained because of its different seasonal meanings in the Southern Ocean (Gregor et al., 2019). To avoid the inclusion of spatial coordinates, the application region, here the ocean model domain, is divided or separated into sub-regions or clusters where physical and biogeochemical processes that drive pCO_2 are relatively coherent and then apply the regression method within each cluster.

In general, data scaling is not necessary for the FNN applications as it has an insignificant impact on the result, but it does help the convergence of the model training and hence, improve the training time (Zeng et

al., 2017). Therefore, to shorten the training time and eventually the reconstruction time, the input variables (v) are normalized, that is, scaled by their mean ($\bar{v}$) and standard deviation (σ_v) as follows: $v = \frac{v - \bar{v}}{\sigma_v}$.

2.5 Framework for regional analytics

For a better understanding of spatial distribution of uncertainties and biases in the Southern Ocean CO2 estimates, a regional comparison framework is derived from the Southern Ocean zones defined by Lovenduski et al. (2008). This thus allows us to gain more insight of results with better analysis and interpretation. The broader region of interest in this study being the Southern Ocean south of 40°S, the regional analytics is based on the zonal definition (Lovenduski et al., 2008) characterized by the Southern Ocean delineation by major fronts including the south subtropical front (SSTF), the sub-Antarctic front (SAF), and the Antarctic polar front (APF) or simply polar front (PF). These fronts represent steep gradients in vertical structure, temperature, salinity, and nutrient concentration (Lovenduski et al., 2008). The climatological locations of the major fronts that are used here are derived from the BIOPERIANT12 model output and calculated as in Orsi et al. (1995). For the purposes of a vigorous evaluation, it is important to assume that the fronts are stationary. As shown in Figure 2.8, the sub-Antarctic (SAZ) is defined as the region bordered by the SAF and SSTF. The polar frontal zone (PFZ) is the region between the SAF and APF/PF, and the Antarctic zone (AZ) is defined as the region south of the APF/PF (Moore et al., 1999; Lovenduski et al., 2008). By definition these three Southern Ocean zones are similar to the components of the regional comparison framewok used by Rödenbeck et al. (2014).

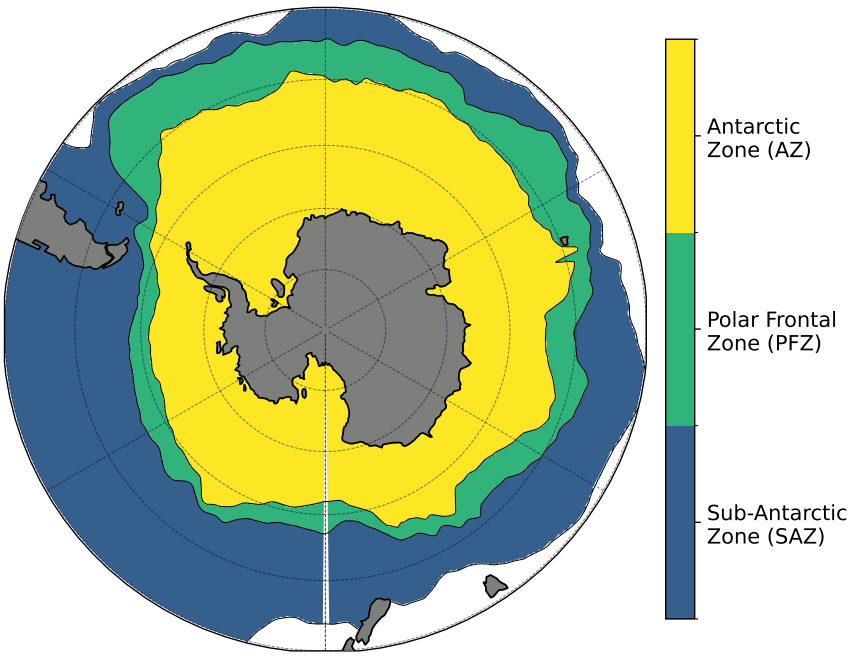
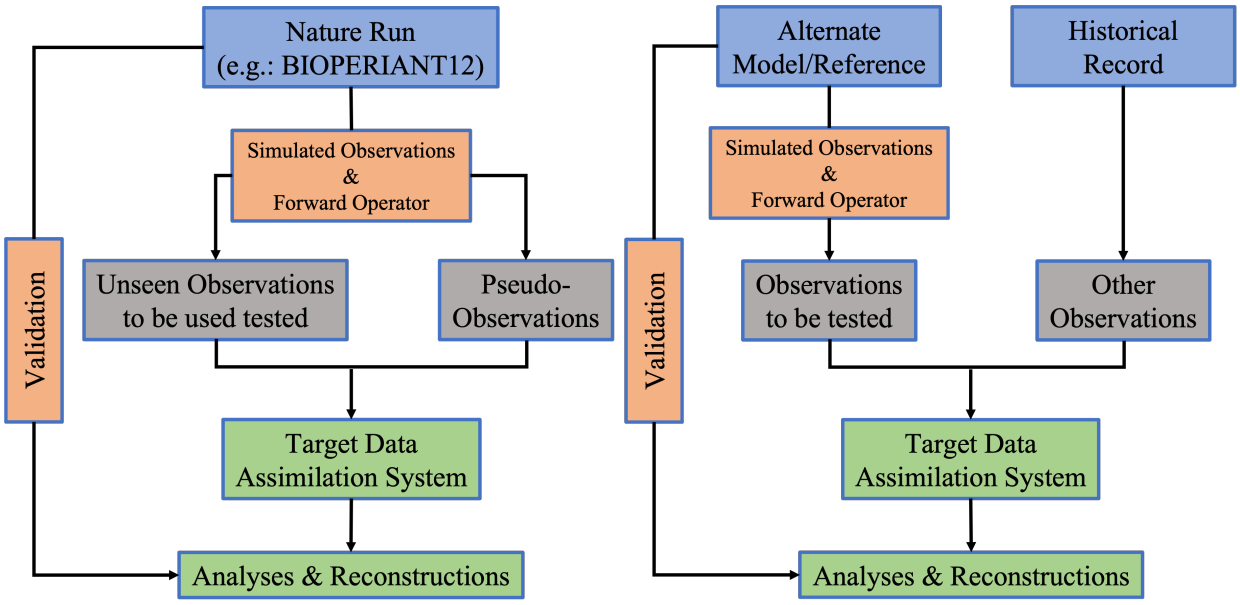


Figure 2.8: The three zonal regions of the Southern Ocean (south of 40°S) based on the zonal definition of Lovenduski et al. (2008). The common names for these zonal regions are shown in the colour bar with their abbreviations in the round brackets.

2.6 OSSEs as tools for evaluating ocean-observing platforms

The observing system simulation experiment (OSSE) is a modelling experiment originally used to assess the usefulness of new observing systems on satellites when actual observations are not available (Zeng et al., 2020; McCarty et al., 2021). In this thesis, I adapt OSSEs for the evaluation of the impacts of existing and new observing systems on operational reconstructions of surface ocean pCO_2 in the Southern Ocean when actual observational data are sparse in time and space.

When planning new observations of surface ocean pCO_2 in the Southern Ocean, that may be used to improve the accuracy of ocean carbon sink and prediction, and subsequently improve our understanding of the climate sensitivity of the Southern Ocean, it is important to assess the likely impacts of the observing systems on the quality of the products. Based on the success of OSSEs in numerical weather predictions, OSSEs can provide us a quantitative means for evaluating ocean observing platforms in the region. This would require developing the ability to simulate ocean sampling platforms using outputs from high-resolution ocean models such as BIOPERIANT12 (Section 2.1).



(a) Adapted traditional OSSE. **(b)** Historical OSSE.

Figure 2.9: Here are two schematic depictions of **(a)** an adapted traditional OSSE (Zeng et al., 2020; McCarty et al., 2021) used in methodology and **(b)** the historical or sampling OSSE methodology currently being explored by the Naval Research Laboratory.

A few studies carried out OSSEs to identify potential geographical locations for making surface ocean pCO_2 measurements using floats (Kamenkovich et al., 2017; Mazloff et al., 2018) or ships and moorings (Majkut et al., 2014; Valsala et al., 2021; Denvil-Sommer et al., 2021). However, a study has not yet been conducted with OSSEs to evaluate the impacts of existing and new observing platforms on operational reconstructions of surface ocean pCO_2. As presented in Section (1.3.1), the two types of OSSEs that are

being used in the community are the traditional and the historical OSSEs (Figure 2.9). While the historical OSSE is currently still being explored by the Naval Research Laboratory (Zeng et al., 2020), throughout the thesis I adopt the implementation of the traditional OSSE whenever necessary (Figure 2.9a).

2.7 Uncertainty decomposition/breakdown

A firm understanding of the uncertainties is required for the purpose of the analysis given that in this 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. I assume that the sampled observations in the OSSEs (Figure 2.9a) are unbiased, and hence the training datasets for surface ocean pCO_2 are considered such as known; and this can be justified by the fact that in a traditional OSSE setting we have access to the whole data (cf. Figure 2.9a). The term errors and uncertainties are used interchangeably 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).

The pCO_2 total uncertainty (E) is dealt with as in (Gregor and Gruber, 2021). The authors identified three main sources of errors that contribute to E within the surface ocean carbonate system. This includes (1) the measurement (M), (2) representation (R), and (3) prediction (P) errors. Under the assumption that these components are independent of each other in the pCO_2 total uncertainty 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:

$$\mathrm{E} = \sqrt{\mathrm{P}^2 + \mathrm{M}^2 + \mathrm{R}^2}. \tag{2.7.1}$$

The contribution of the measurement uncertainty (M) can be removed from this equation (2.7.1) since we are sampling from a synthetic dataset. Further, the representation uncertainty (R) is addressed by sampling at a higher resolution (Gregor and Gruber, 2021). Given that predictions are done at high resolution (daily $\frac{1}{12}$ degree), 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 recommended in Monteiro et al. (2015). Lastly, I assume that machine learning (ML) models are the best possible predictors for the given training datasets, since each ML model is trained using best practices. Therefore, reported root mean square errors will be the uncertainties due to sampling bias.

Chapter 3

Scale-sensitive sampling in pCO_2 reconstructions in the Southern Ocean

This chapter appeared in full as published in the *European Geosciences Union: Biogeosciences* as follows. **Djeutchouang, L. M.**, Chang, N., Gregor, L., Vichi, M., and Monteiro, P. M. S.: The sensitivity of pCO_2 reconstructions to sampling scales across a Southern Ocean sub-domain: a semi-idealized ocean sampling simulation approach, *Biogeosciences*, 19, 4171–4195, https://doi.org/10.5194/bg-19-4171-2022, 2022.

3.1 Introduction

3.1.1 Context and problem

The Southern Ocean remains the world's largest modulator for the ocean uptake of anthropogenic CO_2 (Sabine et al., 2004; Frölicher et al., 2015; Friedlingstein et al., 2021). Therefore, reducing uncertainties and biases in CO_2 budget estimates in the region is important to better assess and understand its influence on regional and global climate (Majkut et al., 2014; Gruber et al., 2019; Hauck et al., 2020). For instance, since the early 2000s, the Southern Ocean 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 (Canadell et al., 2021; LeQuéré et al., 2007). Based on these findings, many studies have been conducted recently to investigate what drives these inter-annual and decadal changes in the Southern Ocean carbon sink and assess the uncertainties of the estimates (Bushinsky et al., 2019; DeVries et al., 2017; Fay et al., 2018; Gregor et al., 2018, 2019; Landschützer et al., 2016; McKinley et al., 2020). However, there have not been many studies looking into the role of intra-seasonal and seasonal modes of variability on the uncertainties and biases reported in empirical CO_2 mapping approaches (Landschützer et al., 2016; Gregor et al., 2019). Novel mapping methods (Landschützer et al., 2016; Gregor et al., 2019, 2017; Denvil-Sommer et al., 2019; Landschützer et al., 2014) 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^{ocn}) 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 the 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 including research and commercial vessels (Bakker et al., 2016; 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 about 2% of all the monthly 1° observational grid points (Bakker et al., 2016; Sabine et al., 2013). Mainly due to its remoteness and harsh weather especially during stormy autumn and winter, it has been increasingly shown that the Southern Ocean is the ocean region that contributes the most to these uncertainties in the contemporary estimates of the mean annual CO_2 uptake (Bushinsky et al., 2019; Gloege et al., 2021; Gregor and Gruber, 2021; Ritter et al., 2017). For instance, sparse observations in largely inaccessible Southern Ocean areas, particularly during the stormy wintertime, have been the biggest barrier to constraining the seasonal cycle of regional and global contemporary ocean-atmosphere CO_2 exchange (Bakker et al., 2016; Monteiro et al., 2015; Ritter et al., 2017; Rödenbeck et al., 2015). However, Gloege et al. (2021) show that the two-step machine leaning approach, SOM-FFN (Landschützer et al., 2014), for pCO_2 recontructions captures the seasonal cycle and amplitude.

Complementary to the increasing effort in the shipboard CO_2 observations through the 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 the Southern Ocean in recent years (Bakker et al., 2016; Bushinsky et al., 2019; Gray et al., 2018; Monteiro et al., 2015). Over the last decade, the advent of a range of new autonomous ocean observing platforms has opened doors toward closing the seasonal and intra-seasonal sampling biases created by the high cost of ship operations in the Southern Ocean outside the summer window (Bushinsky et al., 2019; Gray et al., 2018; Majkut et al., 2014; Monteiro et al., 2015; Sutton et al., 2021; Williams et al., 2017).

Thus resolving the mean seasonal cycle and intra-seasonal mode of variability through in situ observations not only is a challenging exercise but also has followed several avenues from extrapolating findings from the Drake Passage Time-series (like in Fay et al. (2018)) to utilizing measurements from extended deployments of autonomous ocean observing platforms such as Wavegliders (Monteiro et al., 2015; Nicholson et al., 2022), biogeochemical Argo floats (Bushinsky et al., 2019; Gray et al., 2018; Williams et al., 2017), and more recently using Saildrones (Sutton et al., 2021). These advances have allowed the density of the Southern Ocean surface CO_2 observing networks to increase, particularly in the Sub-Antarctic Zone (SAZ) and Polar Frontal Zone (PFZ) which to date are the most observed sub-regions of the Southern Ocean. Consequently, the problem of general sparseness in observations and particularly the sampling biases(Gloege et al., 2021; Monteiro et al., 2015) was partially addressed but not resolved by the ocean CO2 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 pseudo-Lagrangian

platforms that are carried zonally by water currents (Bushinsky et al., 2019; Gray et al., 2018; Majkut et al., 2014; Monteiro et al., 2015; Sutton et al., 2021; Williams et al., 2017). 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). Therefore, an increase in winter sampling would yield a reduction in the uncertainty levels of surface ocean pCO_2 estimates (Bushinsky et al., 2019; Gregor et al., 2019). Notwithstanding these new platforms, sparse and scale-sensitive observations in the Southern Ocean continue to be a barrier to 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 uncertainties and biases underlying pCO_2 reconstructions as reported by Gregor et al. (2019). According to the authors, the performance measures in existing empirical methods converge, which led the authors to the rhetorical question "have we hit the wall?" In practice, high-quality in-situ CO_2 observations like those annually collected and compiled within the SOCAT database (primarily from ships) are fundamental to novel machine learning methods (Bakker et al., 2016; Sabine et al., 2013), despite the reconstructions being limited by spatial and temporal observational gaps and biased sampling (Gregor et al., 2019). As a result, our understanding of the derived impacts of the Southern Ocean dynamics, particularly seasonal and intra-seasonal modes of variability has remained comparatively poor (Gruber et al., 2019), which may have also been contributing to errors in the pCO_2 estimates. At a global scale, (Gloege et al., 2021) coupled an observing system simulation experiment (OSSE) with Earth system models to quantify errors in observation-based reconstructions of air-sea CO_2 exchange by using one of the current gap-filling techniques, the self-organizing map feed-forward neural-network (SOM-FFN) by Landschützer et al. (2016). The authors found that errors were regionally high in the Southern Hemisphere, particularly in the Southern Ocean for which insufficient sampling led to a 31% ($15\% - 58\%$) over-estimation of decadal variability, but they did not discuss the perspective of uncertainties and biases due to intra-seasonal mode of variability.

3.1.2 Aim and question

This study aims to investigate the sensitivity of the pCO_2 reconstructions to the spatio-temporal sampling scales of surface ocean CO_2 observing systems under the assumption that intra-seasonal modes of variability are critical to addressing reconstruction uncertainties and biases. To do that, we used a one-year high-resolution (±10km) coupled physical and biogeochemical forced ocean model for a Southern Ocean sub-region that represents the scales of variability that we aim to resolve. Then, we conduct a series of semi-idealized OSSEs based on existing CO_2 observing platforms (Ships, Wavegliders, Carbon-floats, Saildrones) and coupled with an ensemble of two state-of-the-art machine learning techniques (ML2). A rigorous assess-

ment of the experiment scenarios is conducted through testing and understanding of the ML2 capabilities. We explore the question set by (Gregor et al., 2019) about the prediction uncertainties and biases in contemporary pCO_2 reconstructions being now constrained by the sampling scales achievable by the existing ocean observing platforms. We make proposals toward significantly advancing machine-learning reconstructions "beyond the wall". The goal is to find out how the ocean carbon cycle community can better supplement ship-based observations, essential to pCO_2 reconstructions, with autonomous platform samples in order to reduce the uncertainties and biases of machine-learning-based mapping approaches. The following question will guide this study based on the aim mentioned above:

- Does scale-sensitive sampling matter to empirical reconstructions of the surface ocean pCO2?

3.2 Material and methods

3.2.1 Data processing and derived variables

The data used in this study is from a year-long output of the high-resolution (daily $\times \pm$10km) ocean model, BIOPERIANT12. This ocean model configuration is thoroughly described in 2.1. There are many prognostic variables however, those of particular interest for this study include the coordinates (time, latitude, longitude), the surface ocean pCO_2 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 2.2.

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 MLD and Chl-*a* data that undergo a $\log_{10}$ transformation to return a distribution closer to a normal distribution as presented in Table 2.2. Other advantages of including these variables as proxies are given in Section 2.1.2. It is also substantially beneficial to include only the temporal coordinate as a proxy of pCO_2 because of the characteristics of the study area (Figure 3.1a) as being a single domain with no regional or clustering subsets, otherwise, clustering subsets would be used to overcome the spatial limitations that observations present (Gregor et al., 2019). The inclusion of the time coordinate as a proxy of pCO_2 was done through a variable transformation that aims to preserve the seasonality of the data as presented in Table 2.2.

3.2.2 Study region and selection of the experimental domain

The seasonal cycle is known not only as being the strongest mode of natural variability of CO2 but also the one that most strongly links climate and ocean ecosystems (Mongwe et al., 2018) . Given its characteristics that are largely shaped by higher frequencies such as the intra-seasonal mode of variability defining the re-

sponse modes in physics and biogeochemistry components, the Southern Ocean Seasonal Cycle Experiment (SOSCEx) project was created (see Section 2.2 for more details). As schematically depicted in Figure 2.2, 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 the climate sensitivity of CO_2 and ocean ecosystem dynamics and how these processes are parameterized in forced ocean models such as the regionally configured NEMO-PISCES ocean model, BIOPERIANT12 (Section 2.1.1).

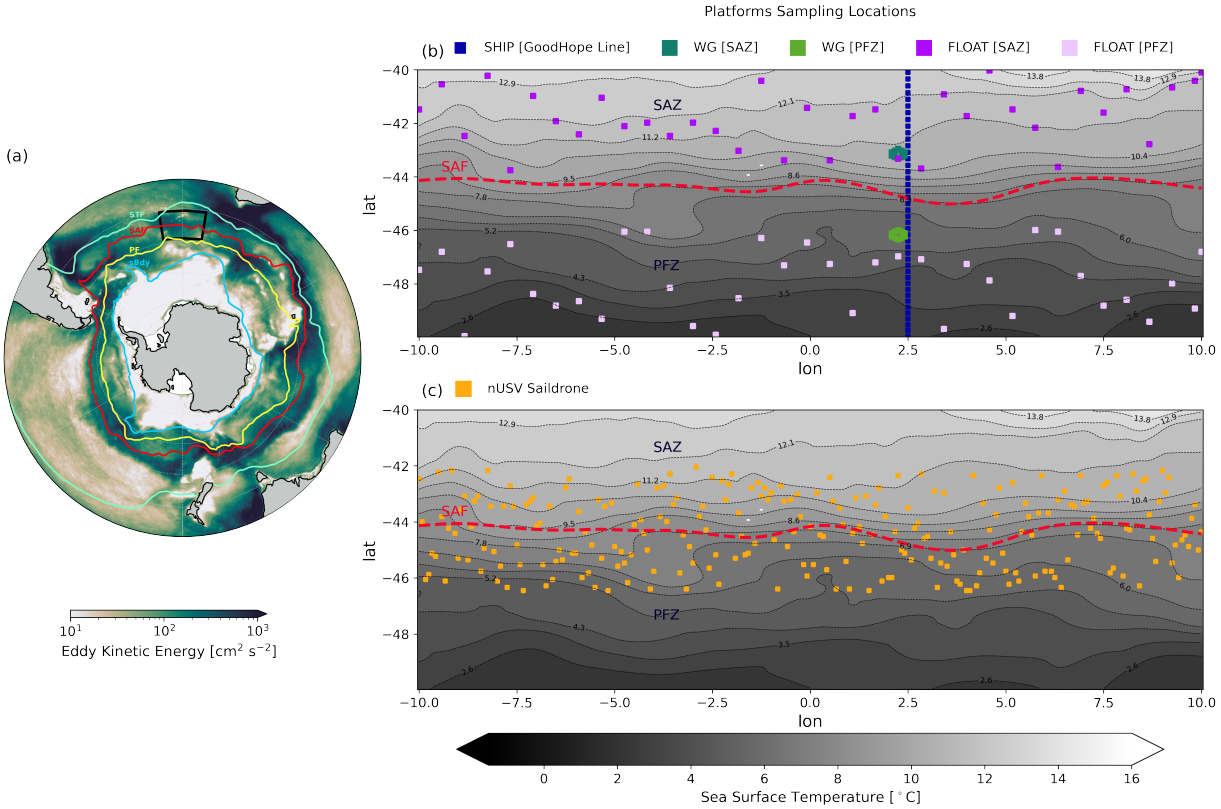


Figure 3.1: Panel **(a)** is the regional view of the BIOPERIANT12 ocean model simulations. It shows the experimental domain (black box) and the annual mean of the Southern Ocean major fronts and the changing conception of the Antarctic Circumpolar Current (ACC), all shown on the map of the mean annual 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)** shows the map of the SST in the experimental domain (black box in Figure 3.1a) 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, characterized by the red line (Figure 3.1a) and red dashed line (Figure 3.1b-c), separates the experimental domain into the Sub-Antarctic Zone (SAZ) and Polar Front Zone (PFZ).

This study is designed as a semi-idealized observing system simulation experiment (OSSE) to minimize some of the potential confounding factors on the final estimation of the root mean square error (RMSE), mean absolute (MAE) and temporal and spatial biases while evaluating the performance of ML regression models used to extrapolate surface ocean pCO_2 values. A key part of this design was to remove the normal

step of clustering whose at large-scale mapping domain is necessary to overcome the spatial and temporal limitations of observations (Fay and McKinley, 2014; Gregor et al., 2019; Landschützer et al., 2014). Thus, to avoid the clustering step, I chose a sub-domain the forced ocean biogeochemical model (BIOPERIANT12) that was not only spatially and temporally coherent but also big enough to reflect the spatial and temporal scales necessary to provide sufficient sensitivity to the different sampling strategies in the Southern Ocean. The selected domain, 10 degrees of latitude (40°S – 50°S) and 20 degrees of longitude (10°W – 10°E) as depicted in Figure 3.1a, is in the Atlantic sector of the Antarctic Circumpolar Current (ACC) between the Subtropical Front (STF) and the Polar Front (PF) and spans across the Sub-Antarctic Front (SAF) (Figure 3.1a). Furthermore, the domain lies within the Sub-Polar Seasonally Stratified (SPSS) biome (Fay and McKinley, 2014). The Good Hope repeat hydrography sampling line passes through the domain (Figure 2.2) 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 crossed by the SAF, therefore, includes the SAZ and the PFZ, inspired by (Gray et al., 2018) and Chapman et al., 2020. The SAZ and PFZ, separated by the SAF (red line Figure 3.1a, and red dashed curve in Figure 3.1b-c), are respectively referred to as the north and south of the experimental domain. The oceanographic context of this domain is shown in Figure 3.1a, depicting the selected $10^\circ - \text{by} - 20^\circ$ domain (black box) in the context of the Southern Ocean major fronts and the eddy kinetic energy (EKE) derived from the BIOPERIANT12 ocean model. This confirms that the domain spans the Sub-Antarctic Front (SAF) and is in a region of relatively high/medium EKE.

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

The mean seasonal cycles of pCO_2 reconstructions from two well-known machine-learning-based products (Landschützer et al., 2016; Gregor et al., 2019) are explored here within the study sub-domain in comparison with the BIOPERIANT12 ocean model pCO_2 (Figure 3.2). In the Southern Ocean, the observed maximum-positive anomaly in surface ocean pCO_2 in winter (Jul - Sep) is linked to mixed layer deepening and associated entrainment, while the maximum-negative anomaly in summer is linked to the spring-summer net primary production (Gregor et al., 2018; Takahashi et al., 2009).

The BP12 model sub-domain (black box, Figure 3.1a) is depicted as a winter-maximum and summer-minimum pCO_2 area by both data products (Gray et al., 2018; Keppler and Landschützer, 2019) as shown in Figure 3.2a-b. Thus, the domain-mean seasonal cycles of pCO_2 from these two products are quite consistent with the broader Southern Ocean (Gregor and Gruber, 2021). This is in sharp contrast with the seasonal cycle climatology from the high-resolution forced ocean model used in this study (Figure 3.1c). 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 (Figure 3.2c). This kind of temperature-driven model bias for surface ocean

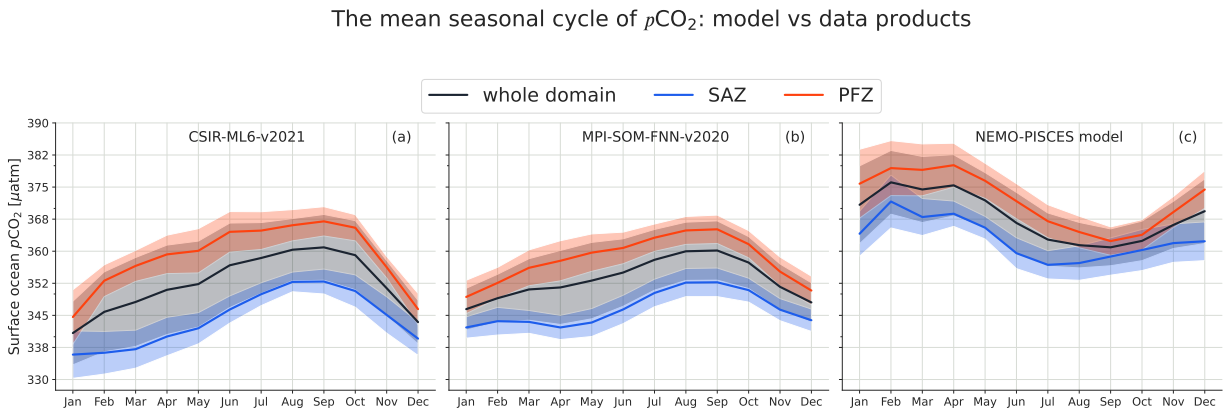


Figure 3.2: The mean seasonal cycle (SC) for surface ocean pCO_2 from two observation-based products and the BIOPERIANT12 ocean model within the selected experimental domain (Figure 3.2a). It contrasts their respective seasonal cycles. Panels **(a)** and **(b)** respectively show the mean SCs of the 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.

pCO_2 is now well recognized in both forced and coupled models in the Southern Ocean (Mongwe et al., 2016, 2018). This study is more focused on the modes of variability within the model than it is with the mechanisms.

The forced coupled ocean model (NEMO-PISCES) represents the processes that regulate CO_2. However, for the purpose of the study, the "correctness" of the pCO_2 response to the driver variables is not really necessary because here we examine the sensitivity of the reconstruction to how the sampling scales match the modes of variability. This work and Gloege et al. (2021)'s study use different ocean models. Gloege et al. (2021) used medium-resolution Earth system models that do not resolve the sub-grid scale dynamics which are explicitly simulated in the BIOPERIANT12 ocean model. As a model resolving the sub-grid scale dynamics in the Southern Ocean, BIOPERIANT12 represents the scales of variability of interest in the study. Thus, the pseudo-observations resulting from sub-sampling the model outputs, taken as ground truth, do reflect the sampling scales of corresponding observing platforms.

3.2.4 Experimental configurations

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 surface ocean pCO_2 or derived (from pH) surface ocean CO_2 observations. More specifically, the existing ocean observing platforms involved in these experiments are the ships (serving as the baseline), 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 (Figure 3.1b-c). The first autonomous platform, Carbon-float, characterizes the autonomous profiling biogeochemical float operating in the Southern Ocean (Majkut et al., 2014; Williams

et al., 2017; Gray et al., 2018). Manufactured by Teledyne/Webb Research or Seabird Electronics, these floats are designed to provide year-round measurements at 10-day periods (Johnson et al., 2017). The second autonomous platform, Waveglider, is an autonomous USV developed by Liquid Robotics Inc (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 (Grare et al., 2021; Sabine et al., 2020). 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 (Gentemann et al., 2020; Meinig et al., 2016, 2019). 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 ocean-observing 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 (Figure 3.1b). 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 is constrained by the SAF (Figure 3.1a-b). Wavegliders were constrained to repeat the pseudo-mooring sampling ($\pm$20 km range) on the ship line (Figure 3.1b), which captures the sub-mesoscale gradients but with a high temporal sampling frequency of 1 hour. Moreover, from a logistic perspective, WGs were given a mooring-like sampling program to ease their deployment and retrieval, for example, from the research vessel SA Agulhas II which crosses the domain at the Good Hope line, whereas nUSVs would be able to sail to the next port.

Idealized experiment setup

In this paragraph, I briefly describe the experimental scenarios shown in Table B.1. I 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. The BIOPERI-ANT12 ocean model simulations is considered as Nature Run in the OSSE; that is, a realistic representation of the real ocean climate systems within which the pCO_2^{ocn} is known across the entire experimental domain. Based on this, the question is: given measurements of pCO_2^{ocn} as sampled in a real-world scenario by these ocean observing platforms, how sensitive are the sampling distribution and resolution to observation-based estimates of pCO_2^{ocn} at every point across the entire experimental domain?

In these experiments, I simulate the sampling tracks/patterns of the synthetic ocean observing platforms SHIP, FLOAT, WG, and nUSV Saildrone (Figure 3.1b-c). In other words, I leverage these synthetic sampling systems to sub-sample the Nature Run inside the experimental domain by constraining the experiment to their realistic and existing counterparts. The Nature Run 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 surface ocean pCO_2 values of the full domain and compared with the original Nature Run field pCO_2^{ocn} to assess the anomalies in reconstructed mean annual and seasonal cycles.

The idealized ship operates according to the sampling scales and strategies of ships involved in the SOCAT collaborative effort. However, here three seasonal sampling regimes are we considered for the underway 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 with 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 Figure 3.1b, even though in reality this might happen due to the occurrence of events such as storms or eddies. Two deployment and sampling scenarios are considered to not disadvantage the floats and to value their large spatial structure: (1) in the SAZ, and (2) in the PFZ (Figure 3.1b). Given the pseudo-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 (Figure 3.1b). The idealized Waveglider operates according to the sampling strategies of the Wavegliders used in the SOSCEx project (cf. Section 2.2 for additional details). Like the idealized float, we considered two deployment stations, the first in the SAZ (cf. Figure 3.1b, hexagonal patterns in dark green) and the second in the PFZ (cf. Figure 3.1b, hexagonal patterns in light green). This idealizes the two deployment scenarios of SOSCEx III gliders (cf. Figure 2.2, hexagonal patterns in blue-yellow) that sampled on an hourly basis. However, given the model temporal resolution that is daily, the idealized Wavegliders samples daily. Lastly, an idealized Saildrone is added to simulate the sampling strategies of its real-world counterpart that can sample for up to 12-month ocean data collection missions (Gentemann et al., 2020; Meinig et al., 2019). As with the idealized Waveglider, the Saildrone also samples daily. Further, we assume that by leveraging its speed the Saildrone sampling can be done across an ocean front, such as the SAF as depicted in Figure 3.1c – a realistic assumption because in reality nUSV Saildrones sample at a much higher frequency (hourly) and can be piloted remotely (Gentemann et al., 2020; Sutton et al., 2021). I assumed that all three autonomous sampling platforms sampled year-round in then experimental domain.

The observing system simulation experiment (OSSE) with nUSV Saildrone is inspired by the study of

Sutton et al. (2021) that used nUSV to sample at a very high resolution and completed in about 6 months the first autonomous circumnavigation of Antarctica providing hourly observations. At this frequency, the nUSV sampling density in this study domain (Figure 3.1c) is realistic due to the size of the sampling domain. Using an extracted subset of the Sutton et al. (2021)-USV dataset within the sub-domain, I find that the Sutton et al. (2021)-USV would take about 16 days to cover the 20°W – E domain, which corresponds to 16 days $\times$ 24 hrs = 384 hourly samples. However, the nUSV sampling pattern (Figure 3.1c) is idealized, with the goal of sampling across the sub-domains on both sides of the front (SAF); that is, in the SAZ and PFZ. By using a back-of-the-envelope approach, I find that the Saildrone would be able to cover the domain in about 45 days using a zig-zag pattern - assuming 42°S to 46°S with each pass covering 2.5°W – E for each pass ($\sim$ 500km) with 8 passes in our domain ($\sim$ 4000km) at a speed of about 2 knots ($\sim$ 3.7km/hr).

In summary, the pCO_2^{ocn} and drivers are sub-sampled using the above-mentioned synthetic sampling platforms, i.e., SHIP, FLOAT, WG, and nUSV Saildrone (Figure 3.1b-c). I 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 I use the ML regression techniques described in Sections 2.3.2 and 2.3.5 to reconstruct the full experimental domain and compare it with the BIOPE12 model truth pCO_2^{ocn} in the full domain to assess the impact of scale-sensitive sampling on reconstructions as anomalies of mean annual and seasonal cycles, which is a key objective of this work.

3.2.5 Machine Learning implementation

A two-member ensemble method (ML2) is used for pCO_2 reconstructions in this study domain. This ML2 approach consists of two state-of-the-art machine learning (ML) approaches: the Feed-forward Neural Network (FNN) and a variant of Gradient Boosting Decision Tree (GBDT) learning frameworks called Gradient Boosting Machines (GBM) as described in Sections 2.3.2 and 2.3.5. The choice of the FNN method is motivated by its recent success in approximating the surface ocean pCO_2 (Denvil-Sommer et al., 2019; Gregor et al., 2019; Landschützer et al., 2016). The choice of the GBDT approach 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 (Gloege et al., 2021; Gregor et al., 2019; Gregor and Gruber, 2021). I use the Scikit-learn and LightGBM Python packages for the implementation of FNN and GBM, respectively. Thus I focus only on the ensemble average ML2 whose stacking process is illustrated in Figure 2.7a. Unlike the two main techniques of reference (Landschützer et al., 2016; Gregor et al., 2019) both of which include a clustering step, in this study I avoided it because of the size of the study domain (Section 3.2.2).

Moreover, given that the observation size in this sub-domain is relatively small, especially for the baseline experiment (SHIP summer-only), immediately split the simulated data into training and testing sets may not capture some key features of the original platform observations. I thus use the entire sampled data for model

building instead of splitting the training data into two sets. To control the overfitting, I incorporate a K-fold cross-validation (CV, with K=4) during the model training in order to find the set of hyper-parameters that enable a better generalization of ML2. The 4-fold CV approach is illustrated in Figure 2.7b. 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 tuning of hyper-parameters is achieved using a Bayes-search CV (BayesSearchCV) instead of a grid-search CV (Section 2.3.7). Further, the assessment of generalization is done through quantitative comparison of the estimates with Nature Run data that were not involved in the simulations of synthetic data. This assessment use the performance metrics presented in Section 2.3.6.

3.3 Results

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

3.3.1 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 ocen model (Figure 3.3a). The meridionally distinct SAZ (north of the domain) ($<$ 368.15 µatm) and PFZ (south of the domain) ($>$ 368 µatm), are separated by the Sub-Antarctic Front (SAF) (Figures 3.1 and 3.3a). This mean map also highlights the importance of mesoscale gradients in both the SAZ and the PFZ domains (Figure 3.3a). 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 Figure 3.3b. It shows that the seasonal cycle of pCO_2^{ocn} is dominated by the influence of the annual cycle of the sea surface temperature (SST) on CO_2 solubility (Mongwe et al., 2016; Munro et al., 2015) with warm late summer (Feb-Apr) and cool late winter (Jul-Sep) (Figure 3.3b). 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 (Figure 3.3b).

Notwithstanding the phasing differences, we still find a comparable winter reconstruction bias in this study (Figures 3.2c and 3.3b) and observation-based products (Figure 3.2a-b). Thus, the question is: is the magnitude of the reconstructed winter pCO_2 maximum 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?

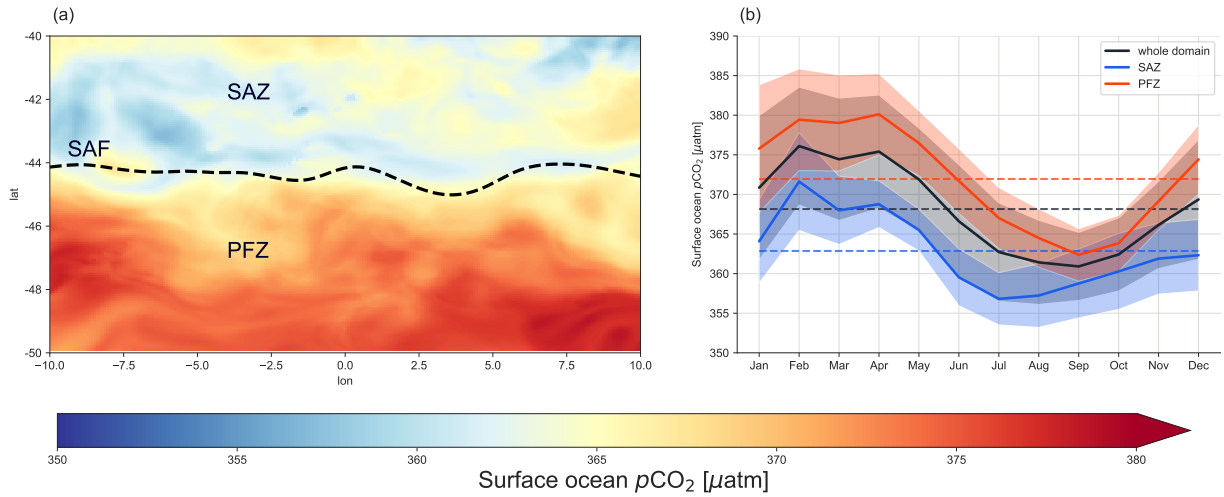


Figure 3.3: Characterisation of the spatial and temporal surface ocean pCO_2 annual mean state within the selected 10°-by-20° experimental domain located in the northern ACC that corresponds to SPSS biome (Fay and McKinley, 2014) as shown in Figure 3.1a. Panel **(a)** shows the map of mean annual pCO_2 from the BIOPERIANT12 (BP12) model. It shows that the domain is characterized by a regional meridional gradient including the Sub-Antarctic Front (SAF) (black 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 in each sub-domain: 368.16 µatm (domain), 362.85 µatm (SAZ), and 371.78 µatm (PFZ).

3.3.2 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.

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 (Figure 3.4b). 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 (Figure 3.4a-b, and Table 3.1), the third one can be found in the Supplementary Assets (Figure B.2) in support of the main points already made in Figure 3.4a-b.

The spatial and seasonal cycle anomalies from the reconstructions for the summer (smr), summer and

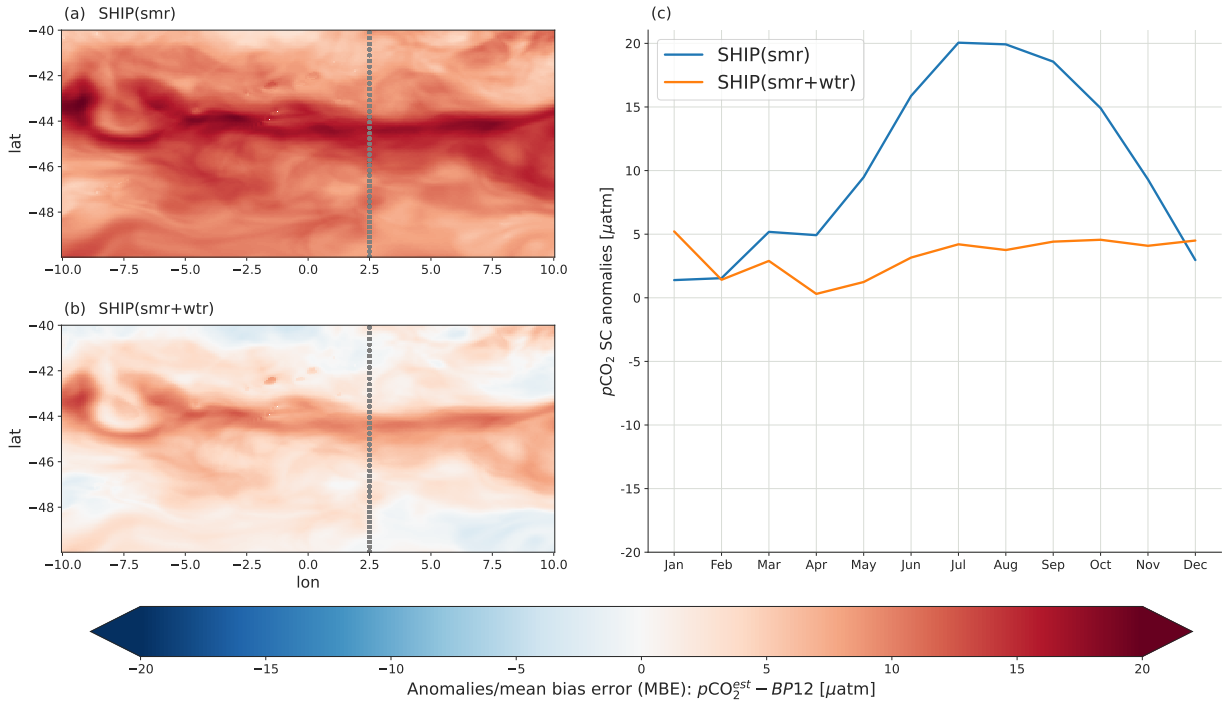


Figure 3.4: Reconstruction anomalies for the SHIP experiment where the idealized ship sampled the domain based on the sampling regimes/scenarios, SHIP(smr) for summer, SHIP(smr+wtr) for summer and winter. Panels **(a)** and **(b)** show the maps of the reconstruction anomalies according to the two sampling regimes SHIP(smr), and SHIP(smr+wtr) respectively; **(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.

winter (smr+wtr) sampling lines are depicted in Figure 3.4a-b. The results for the autumn and spring (aut+spr) sampling lines are summarized in the Supplementary Asset (Figure B.2). The uncertainties and regression errors for all three experiments are shown in Table 3.1. 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 (Figure 3.4a, c). This sampling strategy resulted in a strong positive anomaly ($\pm$20 µatm) that peaks in winter and weakens in mid-summer (Figure 3.4c). 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 (Figure 3.4b-c). The weaker but persistent positive anomaly in the SAF accounts for most of the reduced positive seasonal cycle anomaly (Figure 3.4a, c).

All scenarios depict a mesoscale modulated positive annual pCO_2 anomaly (MBE) climatology in the vicinity of the SAF (Figure 3.4a-b). However, this is slightly offset by equally strong positive anomalies in the SAZ and PFZ for the smr scenario (Figure 3.4a), while the meridional gradients of the anomalies are much weaker for the smr+wtr scenario (Figure 3.4b). These differences are very well reflected in the anomalies of their corresponding seasonal cycles (Figure 3.4c). Table 3.1 shows that in these SHIP-only experiment results, 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 BIOPERIANT12 model ground-truth for pCO_2 ($r = 0.73$).

Table 3.1: 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 these experiments is presented in Table B.1 and described in Section 2.3.4. Similar to Table 3.1, 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, MAE is the mean absolute error; MBE or Bias is the mean average error; and r is the Pearson correlation coefficient between the reconstructed and the BIOPERIANT12 model truth pCO_2. All these metrics are computed following their formulation in Section 2.3.6. Values in the table are significantly different from the mean for the corresponding column (with a 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(smr+wtr)	6.8	5.29	3.18	0.73

Idealized SHIP and autonomous observations platform experiments

In this section, we presented the results of three sets of the 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 (Figures 3.4a, c, and 3.5a-b) (Bushinsky et al., 2019; Gregor 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) (Figures 3.5b, 3.5c-d, and 3.5e-f). These were complemented by simultaneous year-round FLOAT deployments in the SAZ and PFZ (Figure 3.5b, g), and a deployment of the new unmanned surface vehicle (nUSV) Saildrone that spanned across all three domains (Figure 3.5b, h).

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 (Figure 3.5). Statistics (Table 3.2) 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 (Figure 3.5b). 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 (Figure 3.5b). The two experiments with the smallest seasonal biases were the SHIP(smr)

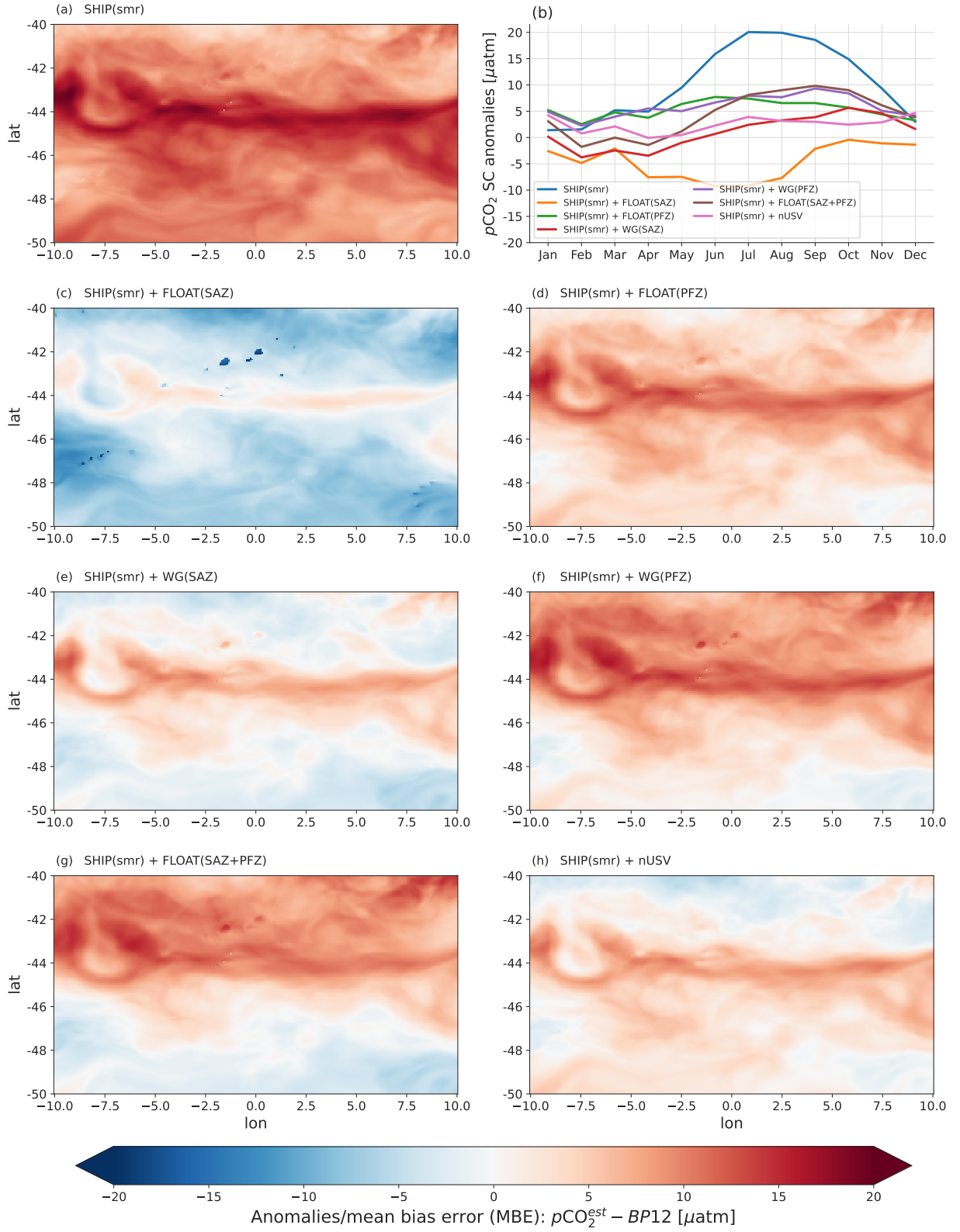


Figure 3.5: Reconstruction anomalies for the SHIP, SHIP + FLOAT, SHIP + WG and SHIP + nUSV experiments with a particular focus on the summer-only baseline scenario: SHIP(smr). **(b)** shows the anomalies of mean seasonal cycle (SC) reconstructions of the summer-only sampling scenario of the above-mentioned sets of experiments, while panels **(a)** and **(c-h)** show the spatial anomalies or biases (MBEs) of the mean annual pCO_2 reconstructions resulting from each these experiments.

+ 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 (4-5 µatm)(Figure 3.5b). 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 ($\pm$10 µatm)(Figure 3.5b).

The spatial annual mean pCO_2 experimental scenario anomalies are consistent with the characteristics of the seasonal cycle of pCO_2 (Figure 3.5a, 3.5c-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 (Figure 3.5a, 3.5c-h). All the scenarios highlight significant mesoscale anomaly gradients across all the domains (Figure 3.5a, 3.5c-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 (Figure 3.5c, e). However, the reverse was found for the SAF zone which shows a stronger positive anomaly for the WG(SAZ) than for the FLOAT(SAZ) (Figure 3.5c, 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 (Figure 3.5b-c). Similarly, the much weaker negative anomalies in the SAZ and PFZ for the WG deployment are consistent with the weaker seasonal cycle ($<\pm$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 (Figure 3.5d, 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) (Figure 3.5b). 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 (Figure 3.5g). The moderately strong positive winter anomalies ($\pm$10 µatm) in the seasonal cycle for this experiment indicate that the mean annual positive anomalies are also dominated by the winter anomalies (Figure 3.5b). The mean annual pCO_2 anomaly for the SHIP(smr) + nUSV deployments is weakly negative ($<$ -5 µatm) in the north SAZ and weakly positive ($<$ 5 µatm) in the SAF and the PFZ (Figure 3.5h). The overall weak mean annual pCO_2 anomaly is consistent with the weakest (0-5 µatm) seasonal cycle anomaly (Figure 3.5b).

Table 3.2 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). This table also shows that although all the additional high-resolution platform experiments reduced the RMSE and MBE, the

Table 3.2: 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 B.1 and described in Section 2.3.4. Any other details for better understanding of the table can be found in the caption of Table 3.1.

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)	8.0	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

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 3.2). 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 (Figure 3.5b), and higher correlations with the known truth (with $r = 0.73$ and $r = 0.74$, respectively).

3.4 Discussions

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 to reduce the uncertainties and biases of the seasonal and mean annual fluxes (Bushinsky et al., 2019; Gregor et al., 2019, 2018; Lenton et al., 2006, 2013; Mongwe et al., 2018; Monteiro et al., 2015; Sutton et al., 2021; Takahashi et al., 2009). Gregor et al. (2019) argued that the uncertainties and biases of pCO_2 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 sea-

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

3.4.1 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 (Figures 3.1b and 3.4a-b, and Table 3.1). 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 $\pm$20 µatm that was strong enough to reverse the in-gassing flux from the model domain (Figure 3.4c), which also biased (positively) the spatial mean annual flux for the domain (Figure 3.4a). The impact of the biased summer sampling is also expressed in the comparatively elevated RMSE: 13.79 µatm (Table 3.1), 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 an average RMSE value of 14.33 μatm in the PFZ, and also RMSE = 13.09 µatm for the SOM-FNN method (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 3.1) 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 $\pm$20 µatmatm in winter to less than 5 μatm over the whole seasonal cycle (Figure 3.4a-c). The impact of the additional winter line is also expressed in the reduction of the bias error (Table 3.1).

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 (Figure 3.6a). 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 (Figure 3.6b).

The splitting 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

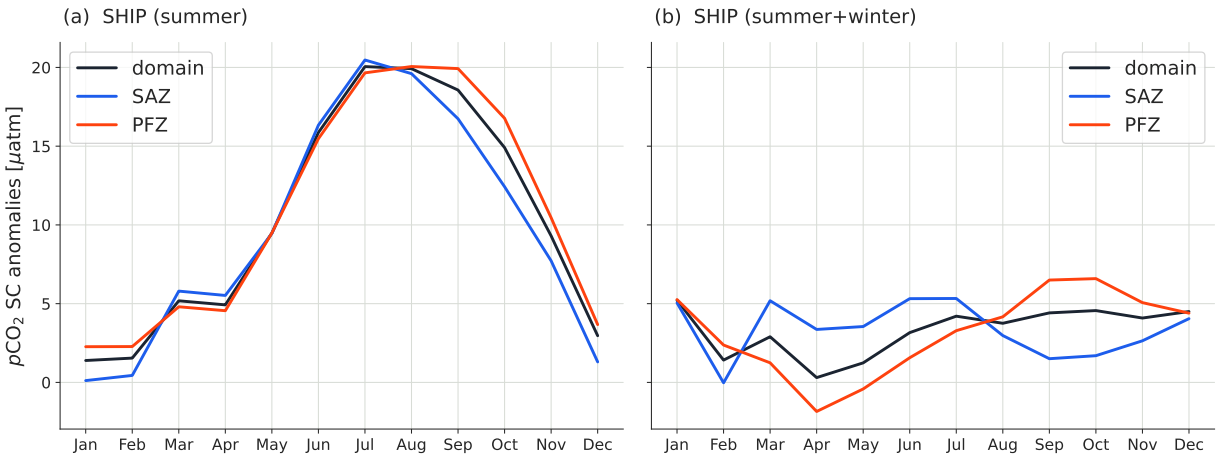


Figure 3.6: Anomalies of the mean surface ocean pCO_2 seasonal cycle (SC) reconstructions from two SHIP-only experiments. **(a)** shows the pCO_2 SC anomalies from the SHIP (summer-only)-based reconstruction in the whole domain, the SAZ, and the PFZ; and in contrast,**(b)** shows the pCO_2 SC anomalies from the SHIP (summer + winter)-based reconstruction for the whole domain, the SAZ, and the PFZ.

as identifying key modes of variability that can influence sampling strategies (Figure 3.6a-b). The 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 30° 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 cost-effective 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 (Figures 3.4a-c and 3.6a-b, and Table 3.1). However, realistically this is not achievable because access to the Southern Ocean outside the summer period is logistically challenging outside the Drake Passage (Gray et al., 2018; Fay et al., 2018; Monteiro et al., 2015).

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-based Carbon Floats, and Saildrones (Bushinsky et al., 2019; Gray et al., 2018; Monteiro et al., 2015; Sutton et al., 2021; Williams et al., 2017). I now discuss the effectiveness of each one through OSSEs simulating their sampling characteristics inside the model domain. All these experiments include the SOCAT-like SHIP-summer observations. The experiments focus primarily on the impact of the autonomous sampling platforms WG and pH-based Carbon Floats as both have been deployed in the Southern Ocean with sampling strategies that view to address the seasonal sampling bias (Gray et al., 2018; Gregor et al., 2019; Monteiro et al., 2015). We return to the potential of Saildrones later in the discussion in the context of a discussion of how to "get over the wall".

3.4.2 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 3.1 and 3.2). Thus, SHIP(smr+wtr) performed better (e.g., RMSE $= 6.88\mu\mu$atm) 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 (Figure 3.7a), or more critically, the seasonal cycle of the meridional gradient of pCO_2 (Figure 3.7b). Together these figures highlight that although the mean increasing southward gradient in pCO_2 is sustained throughout the annual cycle (Figure 3.7a), there are sharp seasonal spatial and temporal contrasts in the meridional variability of the magnitudes (Figure 3.7b). This includes significant seasonal differences in the influence of mesoscale on the spatial variability (Figure 3.7a). 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., 2020; Hauck et al., 2015). 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 3.1 and 3.2) compared to SHIP(smr) + WG and SHIP(smr) + FLOAT.

Significant differences exist between the meridional gradients along the SHIP line in summer (e.g., January) and winter (e.g., July) (Figure 3.7a-b). For example, these differences are more significant farthest south ($> 47°$S) and farthest north ($< 43°$S) compared to the middle $(43°\text{S} - 47°\text{S})$ of the sub-domain (light grey shadings, Figure 3.7a). 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 (Figure 3.7b). 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 (Figure 3.7b). 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

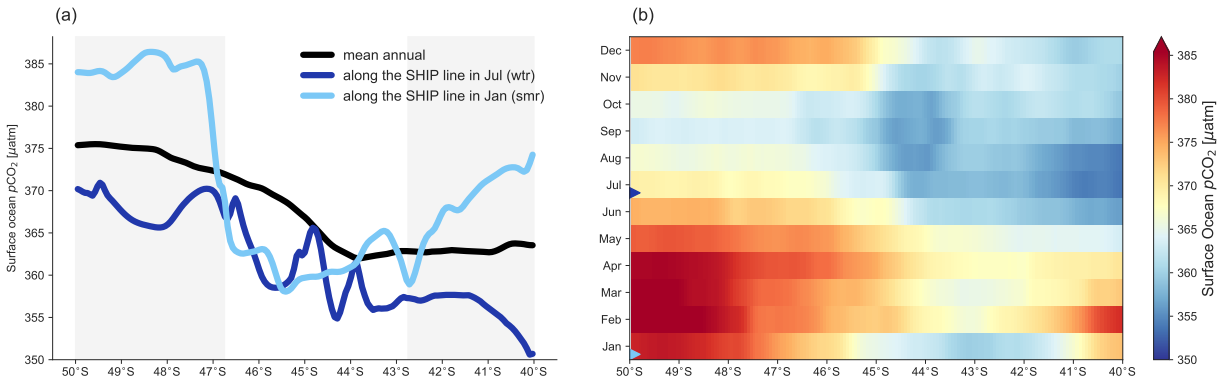


Figure 3.7: Seasonal contrasts for the meridional gradient (MG) of surface ocean pCO_2 in the experimental sub-domain. **(a)** shows the mean annual MG (black), the mean MG along the SHIP line in summer (January)(light blue) and in winter (July)(dark blue); and panel **(b)** shows the seasonal cycle 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.

are supported by the impact that these sampling strategies have on the seasonal cycle of the bias (Figure 3.5b). 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 (Figure 3.5b). 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 (Figure 3.5b, and Table 3.2). 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 (Figure 3.1c). This is consistent with all the instances when not resolving the temporal variability results in a positive bias of varying magnitudes (Figure 3.5b).

On designing an observation-based strategy for quantifying the Southern Ocean uptake of pCO_2, Lenton et al. (2006) argued that constraining the net seasonal air-sea pCO_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.

3.4.3 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 (Bushinsky et al., 2019; Gray et al., 2018; Monteiro et al., 2015; Sutton et al., 2021; Williams et al., 2017). 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 3.2; Figure 3.5). In each case of the FLOAT and WG, they were made to sample each sub-domain (SAZ and PFZ) for a year at their characteristic sampling periods of 10 days and 1 day, respectively. The assumption was that the FLOAT would remain in the domain throughout the year. Thus, to not disadvantage the floats in these experiments, one float was deployed in each sub-domain (SAZ and PFZ) as shown in Figure 3.1b, under the assumption that floats would not cross the Sub-Antarctic Front (SAF). 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 the 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 = 8 µatm; MBE = 5.32 µatm) scenarios (Table 3.2). 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 (Figure 3.8). 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 (Figure 3.8; Table 3.2). 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 (Figure 3.8). 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 (Figure 3.8; Table 3.2).

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. Figure 3.8 then helps explain why even

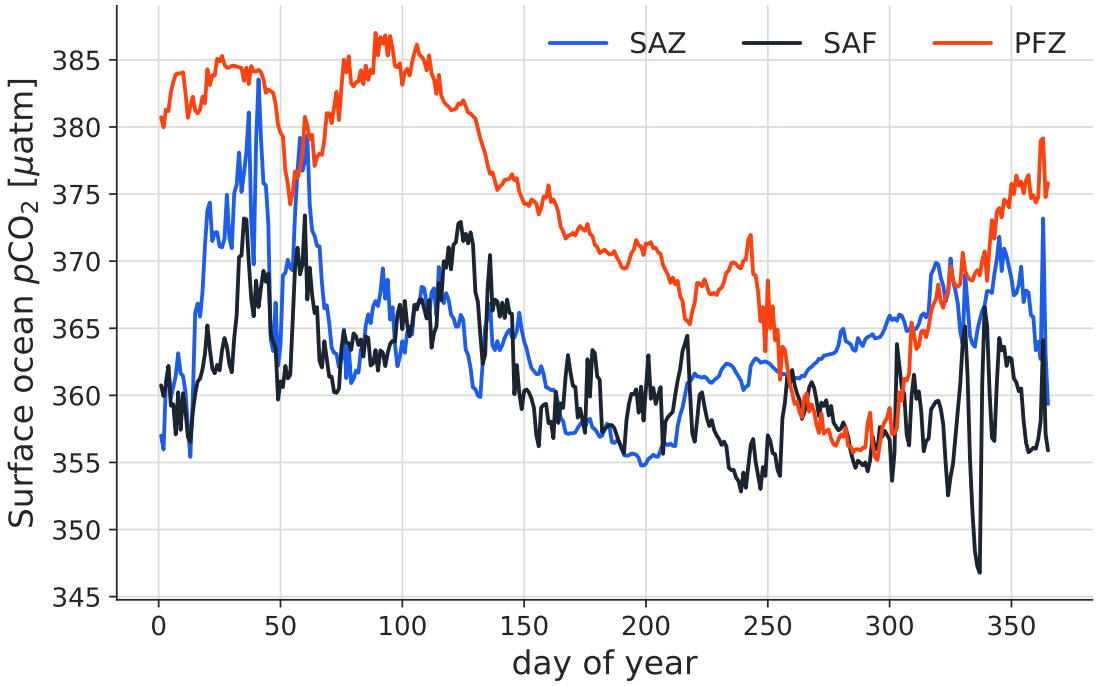


Figure 3.8: Time series (one year) plots of the variability of surface ocean pCO_2 at single model grid cells on the SHIP line (2.5°E, Figure 3.1b). We used the following single model grid cells: 42°S, 2.5°E in the Sub-Antarctic Zone (SAZ); 44°S, 2.5°E on the Sub-Antarctic Front (SAF); and 47°S, 2.5°E 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.

though the domain reconstructions based on the 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. This surprising performance of SHIP(smr) + FLOAT(SAZ+PFZ) after running the experiment several times likely resulted from the difference in modes of variability in the SAZ and PFZ (Figure 3.8). The float did well when deployed in PFZ dominated by seasonal variability which can be resolved by the 10-day sampling period but performed poorly when it was deployed in the SAZ characterized by intra-seasonal modes which cannot be resolved by the 10-day sampling period. Thus, when sampling the two sub-domains simultaneously, SHIP(smr) + FLOAT(SAZ+PFZ) resulted in a poorer performance than for the PFZ alone (Figure 3.5b; Table 3.2). The finding that the high temporal resolution of the SHIP(smr) + WG(SAZ) was the only sampling 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, intra-seasonal and meridional, are close to equally important towards achieving a low bias and RMSE reconstruction. Resolving the former and the latter simultaneously may therefore be a presently missing 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 (Figure 3.5b) and in the uncertainties (32 - 50%), biases (±50%) as well as a statistically significant improvement for the Pearson correlation coefficient (r)

(Figure 3.5a-b, and Table 3.2). 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) + 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 of 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). Based on hourly observations of the surface ocean 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, 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 its 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 (Bushinsky et al., 2019; Gray et al., 2018). This study suggests that these 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. This 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 . It shows that the intra-seasonal modes of the wind are not sufficient to impart a low mean annual and seasonal cycle bias (Bushinsky et al., 2019).

To provide a more quantitative characterization of these findings, an additional analysis was conducted on the sub-10-day mode of variability. A 10-day rolling mean was used to eliminate or weaken the sub-10-day mode of variability (Figure B.1a). The difference between this 10-day rolling mean and the daily model output gives the high-frequency variability and the root-mean-squared error (RMSE) gives us a statistical understanding of what the uncertainty might be if we sampled at a 10-day rate (shown in Figure B.1b as a map). The resulting mean RMSEs for the SAZ and PFZ, after implementing the 10-day rolling mean, are 2.53 µatm and 1.71 µatm respectively, a significant reduction relative to the RMSEs for the FLOAT experiment using the daily model output (Table 3.2). This provides further quantitative support for thes findings and the work of (Monteiro et al., 2015) that more dynamic regions require higher sampling rates. We finally propose that the impact of SOCCOM floats on the reconstructions can be strengthened by reducing the sampling period to < 2 days, especially in high EKE areas, and through a coordinated meridional deploy-

ment strategy 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. This aspect is now examined in more detail.

3.4.4 A proposed optimal sampling strategy for "getting over the wall"

In this study, the 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, I designed an additional year-round Observing System Simulation Experiment (OSSE) that simulated the spatial and temporal sampling capabilities of the new unmanned surface vehicle (nUSV) Saildrone (Sutton et al., 2021) to supplement the SHIP summer-only sampling SHIP(smr) (Figures 3.1c and 3.5b, h); that is, SHIP(smr) + nUSV. This experiment combined the speed of the nUSV Saildrone (Gentemann et al., 2020; Meinig et al., 2015) required to cover the regional meridional spatial gradients length scales (Figure 3.1c), with high-frequency daily sampling to supplement SHIP(smr). Together these fulfill the requirements that emerged from the earlier analysis.

Comparative statistics show that the SHIP(smr) + nUSV experiment yielded a very significant improvement in the reconstruction skills relative to all other platform combinations (Table 3.2). 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 minimizing uncertainties and bias in pCO_2 reconstructions. Based on 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 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 (Figure 3.1b) 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 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

relative to SOCAT-based reconstructions which, as discussed earlier, is not where autonomous sampling vehicles can add the best value (Tables 3.1 and 3.2). These findings 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 high confidence in the pCO_2, and inferred CO2 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.

3.4.5 Applicability of the sub-domain to a wider Southern Ocean

The focus of this study was on investigating the mismatch between sampling periods and the modes of variability of pCO_2 in the domain rather than the mechanisms. This selected domain in the South-East (SE) Atlantic Ocean encapsulates the contrasts in the scales of variability of interest, namely the seasonal and intra-seasonal modes that are characteristics of the Southern Ocean (Figure 3.9). It shows how findings in the study domain can be extended to the Southern Ocean. Using a 10-year period of pCO_2 output from NEMO-PISCES model simulations at a 5-day temporal mean, the Seasonal Cycle Reproducibility (SCR) of pCO_2 was calculated as the correlation of the detrended pCO_2 with its own 10-year climatology – the larger the correlation, the stronger the SCR (Thomalla et al., 2011). This resulted in the SCR-based clustering of the Southern Ocean into three regions (Figure 3.9) corresponding to the low (LSCR), medium (MSCR), and high (HSCR) SCR areas, respectively. The criteria of the choice of these three ranges are as follows. In high SCR areas, there is no intra-seasonal variability and annual signals. In medium SCR areas, intra-seasonal variability emerges but is smaller in magnitude compared to the seasonal cycle, while in low SCR areas, there is no seasonal signal, and the intra-seasonal variability is larger than the seasonal cycle.

Although this study domain was chosen within a high-EKE area (black box; Figure 3.1a; Figure 3.9) because of its contrasting seasonal and intra-seasonal variability of the surface ocean pCO_2, the SCR metric shows how the study area in the SE Atlantic Ocean contrasts to the Southern Ocean as a whole (Figure 3.9). As argued in the previous paragraph, seasonal and intra-seasonal variability is relatively associated with LSCR ($0-0.65$) and MSCR ($0.65-0.85$) regions, which together represent $\sim$75% of the study domain, and $\sim$64% of the whole Southern Ocean (cf. table shown in Figure 3.9). This demonstrates that the subdomain modes of variability (which are dominantly intra-seasonal) may be applied to the wider Southern Ocean. Futhermore, longitudinally, the Southern Ocean is equal to $360^\circ/20^\circ = 18$ times the $20^\circ\text{W}-\text{E}$ sub-domain (Figure 3.1a). While in theory, this sub-domain is 1/18th of the zonal extent of the Southern Ocean, it represents different modes of variability as argued above. Thus, less than 18 USV Saildrones should be able to capture the variability in the whole Southern Ocean.

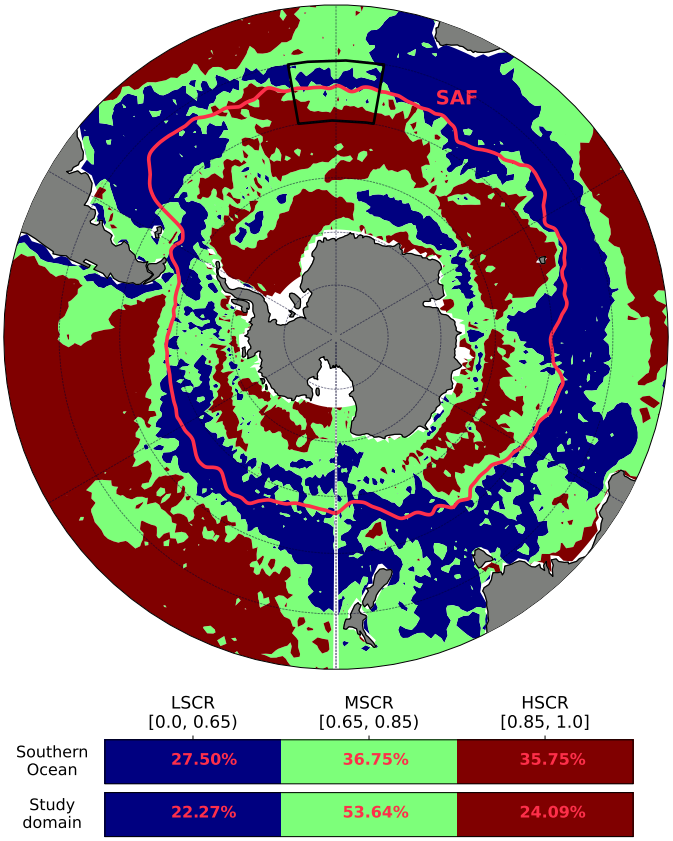


Figure 3.9: Map showing the study domain and the Southern Ocean sub-regions resulting from the seasonal cycle reproducibility (SCR) of pCO_2 calculated based on 10 years of NEMO-PISCES simulations at the 5-day temporal resolution, where the Sub-Antarctic Front (SAF) (light red) and the study domain (black box) are depicted. The table below the map shows the fraction coverage estimates (in %) for these SCR-based regions both in the study domain and Southern Ocean as a whole. LSCR corresponds to low SCR areas, while MSCR and HSCR respectively represent medium and high SCR areas.

3.5 Conclusions

From this study, we propose that one can advance the uncertainties and biases from machine-learning pCO_2 reconstructions "beyond the wall", at least in the Southern Ocean. Within a chosen experimental domain of the Southern Ocean, we demonstrate that this would require resolving the seasonal and intra-seasonal modes of variability of the meridional gradients of pCO_2 through a combination of high frequency (at least daily) observations spanning the meridional axis. We showed that the reconstructed seasonal cycle anomaly and mean annual pCO_2 are highly sensitive to seasonal sampling biases. The 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 domain, which likely may also be contributing to the apparent winter-maximum outgassing or weakening of the ingassing of CO_2 observed in recent Southern Ocean data products. This points to an urgent need to address the existing seasonal bias (towards summer) in the Southern Ocean SOCAT dataset through improving the sampling strategy of the present autonomous platforms, so they are better aligned to the integrated spatial and temporal sampling

scale needs.

Inside the chosen domain, the study confirmed that not resolving the high frequency (synoptic - sub-seasonal) variability results in insufficient decreases in mean biases and RMSE scores for the reconstructed mean annual flux. Present 10-day sampling periods of floats have a limited impact on reducing uncertainties and biases in pCO_2 mappings because they do not resolve the intra-seasonal variability. In addition, the predominantly zonal and quasi-Lagrangian sampling does not contribute sufficiently to resolving the seasonal variability of the meridional gradients of pCO_2. Our study proposes that a more meridionally coordinated deployment of floats could contribute further to resolving synoptic variability and the meridional gradients. For example, increasing sampling frequency to < 2 days, particularly in high-EKE areas as well as a meridionally coherent sampling strategy would support resolving the synoptic-scale variability and the variability of the basin-scale gradients. Although they still lack the meridional gradient reach, Wavegliders in pseudo-mooring modes improve on floats (RMSEs, MBEs) and the main explanation for this improvement is because of their higher sampling frequency (daily). This study recommends that the use of Wavegliders in the reconstruction of CO_2 fluxes in the pseudo-mooring mode should 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 might not be the most efficient way to maximize their strengths stemming from their high sampling frequency (hourly) and large spatial scale (by leveraging their speed). We thus propose that USV Saildrones are probably the optimal platforms to address the necessary integrated large-scale spatial and high-resolution temporal sampling.

In summary, ship-based observations (SOCAT-like) remain vital to the reconstruction of CO_2 fluxes in the Southern Ocean as a whole and should be continued. These observations are the baseline data involved in the training of any machine learning algorithms behind the main observation-based products of reference. However, these ship-based observations are seasonally biased (towards summer) due to under-sampling during stormy autumn and winter seasons, which are likely the root of persistently elevated uncertainties and a winter-positive bias in the reconstructions. This bias should be addressed with urgency. Finally, this study proposes that a meridional sampling strategy may be an efficient way of sampling using autonomous observing systems. In this case, we recommend that existing ship-based observations of the surface ocean pCO_2 in the Southern Ocean should be supplemented by year-round autonomous high-resolution observations that resolve the seasonal cycle of the meridional gradients of the surface ocean pCO_2. However, a follow-up study is also recommended to test, for example, the USV Saildrone effectiveness and impact on reducing uncertainties and biases of the seasonal and mean annual reconstruction of CO_2 fluxes in the Southern Ocean as a whole.

Chapter 4

Sensitivity of two-step pCO_2 reconstructions to clustering choices for the Southern Ocean as a whole

4.1 Introduction

4.1.1 Context and problem

In recent years, two-step empirical modelling and/or semi-supervised machine learning (ML) approaches have been widely adopted in reconstructions of the surface ocean pCO_2 (Landschützer et al., 2016; Landschützer et al., 2014; Gregor et al., 2019; Watson et al., 2020; Gregor and Gruber, 2021; Denvil-Sommer et al., 2021). By two-step empirical modelling and/or semi-supervised ML, we mean a two-step technique in which the first step consists of clustering or applying an unsupervised ML method to the underlying reconstruction domain, and the second step consists of applying a supervised ML method. For example, both Landschützer et al. (2014) and Gregor et al. (2019) adopted a two-step machine-learning approach by first dividing the global ocean into clusters or regions, and then they used a machine-learning regression algorithm to map the surface ocean pCO_2 within every generated cluster.

High activities of eddy kinetic energy (EKE) and the large zonal spatial gaps with high dynamic variability between the meridional lines in the Southern Ocean are a significant challenge in pCO2 reconstructions in the region (Landschützer et al., 2014; Gregor et al., 2019). The community has addressed this problem by adopting different clustering techniques as the initial step of the two-step reconstruction approach using machine learning techniques, but it is not clear what each of these clustering choices contributes to the root mean square error (RMSE) score and bias.

Essential to these machine learning techniques are high-quality in-situ pCO_2 observations (Hauck et al., 2020), such as those annually compiled within the Surface Ocean CO_2 ATlas (SOCAT) through international collaborative efforts (Bakker et al., 2016, 2020). These observations are primarily from ships (Bakker et al., 2016) and constitute the baseline data in any observation-based reconstructions of surface ocean pCO_2 using machine-learning techniques (Hauck et al., 2020; Friedlingstein et al., 2021). Therefore, the existing disagreements in methods in the Southern Ocean are likely primarily driven by the sparseness and seasonally biased distribution of surface CO_2 observations in the region (Rödenbeck et al., 2015; Gray et al., 2018; Gregor et al., 2018; Bakker et al., 2016; Bushinsky et al., 2019; Watson et al., 2020). This raises the need

to explore the sensitivity of the pCO_2 reconstruction uncertainties and biases to the choice of the clustering method in the Southern Ocean. This could not be done in Chapter 3 because of the size of the study domain (Section 3.2.2) for which it was not necessary to use a clustering step in the reconstruction.

4.1.2 Aims and Questions

The primary aim of this chapter is to evaluate the clustering methods, the first step in the two-step pCO_2 reconstructions and the contribution that each of the clustering methods or their ensemble make to the bias and root mean square error. In order to do this, I use a series of observing system simulation experiments (OSSEs) for which I choose to use both the underway SOCAT ship data (pseudo-observations) which is biased toward summer (SOCAT-only) due to very few observations in the Southern Ocean during winter, as well as the SOCAT-only data with a winter sampling gap mitigation (SOCAT + WGM). This choice is motivated by Chapter 3's findings, but now applying them to the whole Southern Ocean (Section 3.4.5). Both OSSE sampling scenarios (SOCAT-only and SOCAT + WGM, detailed in Section 4.2.2) were compared in the evaluation of the clustering methods. Thus, the two questions that guide this chapter towards its aims are as follows.

1. How sensitive are the two-step pCO_2 reconstructions to differences in assumptions that define the clustering step of the methods used in terms of their contribution to biases and uncertainties? In other words, which one of the clustering methods used offers greater gains in terms of reducing RMSE score and bias?

2. How is the contribution of these clustering methods to uncertainties of the reconstruction sensitive to the contrasting summer and winter sampling scenarios?

4.2 Methodology

The two-member ensemble method, including the feed-forward neural network (FNN) and gradient boosting machines (GBM), which was used in Chapter 3 revealed that the GBM algorithm was highly more susceptible to overestimating the surface ocean pCO_2 than the FNN algorithm. This tendency of the GBM algorithm to overfit the training data is due to its potential for high complexity (Frery et al., 2017). Henceforth, only the FNN algorithm is used for the regression modelling of pCO_2, and its technical details and theory are presented in Section 2.3.2.

In this chapter, three two-step methods are explored for the reconstruction of surface ocean pCO_2 in the Southern Ocean. These two-step methods only differ from their first step or clustering step. This step consists of clustering or dividing the pCO_2 domain, here the Southern Ocean, into clusters or sub-regions based on three different approaches including (1) the regionally fixed biomes resulting from the annual

mean of Fay and McKinley (2014) biomes (FMB), (2) the seasonal cycle reproducibility (SCR) of pCO_2, and (3) the mini-batch K-Means (MKM) clustering algorithm. The second step or regression step in all three approaches uses a similar machine learning algorithm (FNN). Given that these two-step approaches use all the same machine learning algorithm for regression, to limit redundancy, when necessary, the resulting two-step gap-filling methods will often be referred to as simply FMB, SCR, and MKM instead of FMB-FNN, SCR-FNN, and MKM-FNN. While details about the data sources and the theory concerning the implicit second step (i.e., FNN regression) of these learning approaches have already been introduced in Section 2.3.2, details of the clustering step of these methods will further be given in this methodology section.

Methodologically, there is an important way the clustering step's influence on errors in observation-based estimates of pCO_2 can be assessed in the Southern Ocean: by improving the temporal availability of ship measurements using an observing system simulation experiment (OSSE) based on a synthetic truth from high-resolution ocean model output introduced in Section 2.1. This methodological section is thus organized around four main features (but not restricted) – which characterize the OSSE framework – including (1) data selection and preparation, (2) platform data simulation, (3) the Reconstruction System and (4) Forecasting/reconstruction and comparison with the ground truth also called Nature Run.

4.2.1 Data selection and preparation

In the supervised machine learning workflow, data are divided into two groups: one to train the underlying algorithm, called the training data, and the other to evaluate the resulting model, called the prediction data. The target and proxy variables data are selected from the daily outputs of one year of the BIOPERIANT12 ocean model (see Section 2.1) which is the high-resolution ($\pm$10 km) ocean model resulting from the Southern Ocean configuration of the NEMOS-PISCES forced ocean model used throughout this thesis. The motivations of the choice of this model is presented in Section 2.1.

For the SOCAT-ship tracks simulation, the 2010-2019 period is chosen in the 2020-release SOCAT data. The motivation behind this choice is that I found that since 1970, over 46% of sea surface CO_2 observations – from this version of the SOCAT database – in the study domain (which is the Southern Ocean south of 40°S) were made between 2010 and 2019. This results from increasing efforts of volunteering ships in international contribution to the SOCAT database (Bakker et al., 2016). I thus assume that in that decade the geographical coverage of annual ship voyages to the Southern Ocean has stabilized. The purpose of the SOCAT dataset is to get the geographic constraints for each line and not just the sampling locations. So, even though a particular line may be sampled annually with 10 repeats, the daily outputs of the BIOPERIANT12 ocean model, being a single year is only sampled once. Together, all the lines create a "synthetic" year which is then disaggregated into summer and winter lines (Figure 4.1).

More specifically, from the 2010-2019 SOCAT data (2020 release) in the Southern Ocean (south of

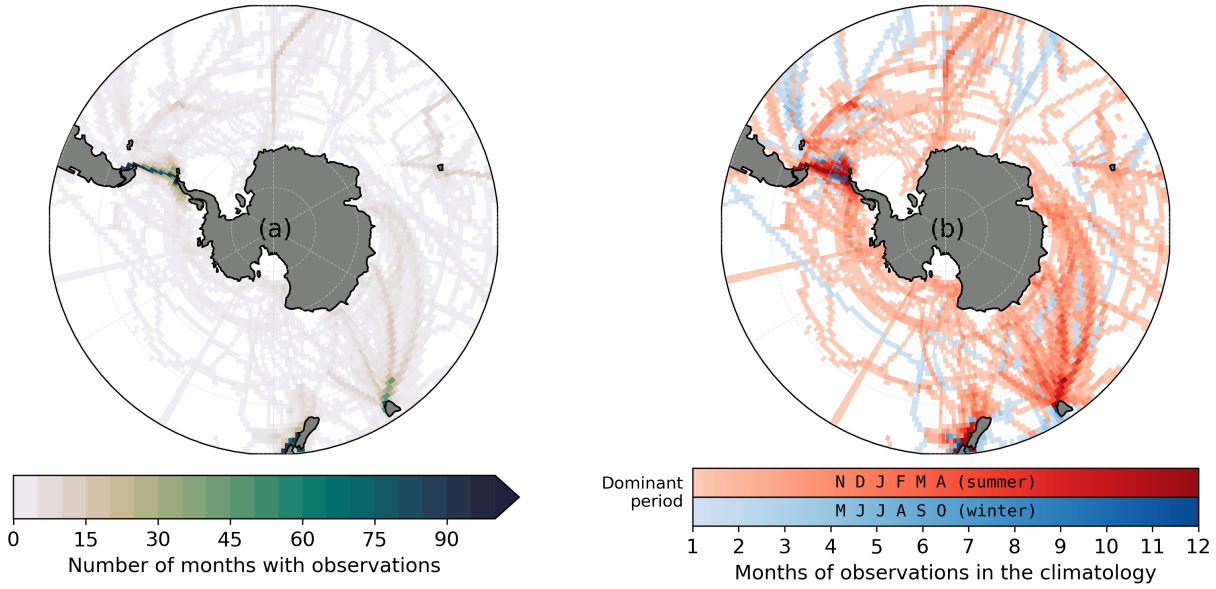


Figure 4.1: Map showing the observation density of SOCAT (version 7) monthly gridded product (2010-2019). Regardless of the year, panel **(a)** shows the number of months observations happened. Panel **(b)** is a monthly climatology to show how the seasonal cycle is represented regardless of the year. The shading in red shows the grid points where the dominant observational period is from November to April (i.e., the big summer), and the blue shading shows the grid points where the dominant observational period is from May to October (i.e., the big winter).

40°S), I extracted the corresponding ship lines in both summer and winter. The extraction is done in such a way that if a ship had sampled 10 times along a line during that period, it will be implemented once for the single model year. Similarly, if sampled only once within the 10-year period, this line is included in the ship lines (Figure 4.1) used to sub-sample the single-year ocean model outputs. Thus, the ship tracks for summer and winter sampling (see Figure 4.1) are used to produce the ship-based synthetic year of pseudo-observations from the model sub-sampling.

Moreover, in Chapter 3, it was found that to reduce uncertainties in pCO_2 reconstructions it is critical to resolve two sampling scales including the seasonal cycle of the meridional gradients (first-order requirement) and the intra-seasonal variability (second-order requirement), which Chapters 4 and 5 focus on. Therefore, to carry out the semi-idealized sampling experiment to the Southern Ocean (south of 40°S), the sampling assumption is that the extracted ship lines could all be sampled during a single year in the region. This approach will likely reduce the bias of the SOCAT-only simulation slightly (Section 4.2.2), but it is assumed not by much as most additional lines are in summer (Figure 4.1) and would have contributed little to improving the uncertainties and biases.

4.2.2 SOCAT data sampling simulation

Figure 4.1 depicts the ship tracks/routes from the 2020 release of the SOCAT database (version 7) in the Southern Ocean where the observation density in the region is not homogeneous (Figure 4.1), with the known high-latitudes under-sampling in time and space (Monteiro et al., 2010; Gray et al., 2018). These SOCAT-

ship routes are extracted and used to set up a series of observing system simulation experiments (OSSEs) that simulate the way ships sample in the Southern Ocean based on two sampling simulation scenarios including SOCAT-only and SOCAT + Winter Gap Mitigation (WGM). Details of these two experiments or OSSE scenarios (SOCAT-only and SOCAT + WGM) are given below.

SOCAT-only sampling data simulation

The SOCAT-only scenario represents the typical SOCAT data in the Southern Ocean, which is biased toward summer with large winter sampling gaps as shown in Figure 4.1b. The SOCAT-ship tracks are extracted and coupled with the BIOPERIANT12 model outputs to simulate SOCAT shipboard measurements of the surface ocean pCO_2 and proxies. The model outputs cover well in space and time the real sampling locations of the SOCAT underway observing ships. Therefore, simulated data resulting from this sampling scenario corresponds to model data that co-localises geographically and temporally the locations of SOCAT shipboard measurements in the Southern Ocean.

SOCAT + WGM sampling data simulation

SOCAT + WGM corresponds to the second sampling simulation scenario of the OSSEs and consists of mitigating the winter observational gaps that make the simulated data from the SOCAT-only scenario biased toward summer. In this scenario, the assumption made is that the same underway SOCAT ship tracks in summer are visited in winter to fill the observational gaps, hence the WGM component (winter gap mitigation) in the SOCAT + WGM scenario which is an extreme sampling case in the Southern Ocean.

In these SOCAT data simulations, I consider only ship observational routes that cover the period from 2010 through 2019 with at least one in-situ observation of surface ocean pCO_2 per grid cell (Figure 4.1). As presented in Section 4.2.1, my motivation is that during this decade, the largest increase in the density of surface ocean CO_2 observations in the Southern Ocean (south of 40°S) was observed. Given that the Nature Run covers very well in space and time the real sampling locations of SOCAT cruises in the Southern Ocean (Figure 4.2.2), the model outputs co-localizing with these geographical and time locations will be called pseudo-observations, which will be used as the training dataset.

The resulting pseudo-observations from both SOCAT-only and SOCAT + WGM scenarios are used as training datasets for the underlying machine-learning method in the OSSE framework introduced in Section 4.2.3. In the SOCAT + WGM scenario, the SOCAT data (biased toward summer) is supplemented by pseudo-observations from synthetic measurements carried in winter along the underway SOCAT ship routes in the Southern Ocean. This OSSE scenario is thus essential to achieve the primary aim of this study (Section 4.1.2).

4.2.3 OSSE setup: experimental overview

The BIOPERIANT12 ocean model output described in Section 2.1 is termed Nature Run in this setup. This simulation is intended to be the best possible representation possible of the true environmental process at higher spatial ($\pm$10km) and temporal (daily) resolutions, which are considered sufficient for testing the observing systems. The framework of the observing system simulation experiments (OSSEs) including the underlying two-step reconstruction approach is presented below and briefly summarized in Figure 4.2 showing a schematic illustration of this experimental design or OSSE framework. A broad overview of the OSSE type this framework is inspired from is presented in Section 2.9.

1. The first stage denoted as Data Simulation consists of selecting from Nature Run the pseudo- or synthetic observations in the Southern Ocean (Section 4.2.2) using existing underway ship routes as it is in the SOCAT database (2020 release).

2. The second stage denoted as "Southern Ocean clustering", consists of dividing the Nature Run field (Southern Ocean) into clusters or regions having some specific physical and biogeochemical properties. Here, the Nature Run field is divided into clusters or sub-regions using a clustering technique. Three different clustering approaches (details are in Section 4.2.4) including (1) Fay and McKinley (2014) biomes (FMB), (2) the seasonal cycle reproducibility (SCR), and (3) the mini-batch K-Means (MKM).

3. The third stage consists of a Reconstruction System. Here, synthetic observations are assimilated into various configurations of the operational machine-learning model. Specifically, the Reconstruction System setting is characterized by the second step in the two-step reconstruction method (detailed in Section 4.2.4) as follows.

   (a) For each clustering method or configuration (FMB, SCR, MKM), a machine learning regression algorithm, the feed-forward neural network (FNN), is applied to the resulting clusters of the Southern Ocean, which results in 4 FNN models using FMB-based clusters (4), 6 FNN models using SCR-based clusters (6), and 7 FNN models using MKM-based clusters (7). Overall, this makes up a total of 17 FNN models implemented, resulting from the three two-step-learning methods FMB, SCR, and MKM.

   (b) In each cluster, an interactive train-validation approach (rather than just one validation split) is used to develop an FNN model on pseudo-observations (synthetic SOCAT data) within that cluster. This Reconstruction System component, denoted by "hyper-parameter tuning", leverages the Bayesian optimization and K-fold cross-validation (CV) (see Section 2.3.7 for more detail on model tuning), implemented as `BayesSearchCV` in the `Python` package `scikit-optimize`.

4. Forecasts or reconstructions of surface ocean pCO_2 are made with the optimal model resulting from the step above. Afterwards, a thorough analysis of the results and comparison of the FMB, SCR, and MKM methods using the test data, which corresponds to the remaining Nature Run that was not used to back out synthetic observations. This allows not only the testing of the reconstruction power or forecast ability of these three methods but also the quantification of improvements due to the consideration of new sampling regimes of the observing system (ship).

4.2.4 Two-step reconstruction methodology

First step: clustering choice in the Southern Ocean

The three clustering approaches introduced early are further described to create regions/clusters/biomes of internal coherence with respect to the surface ocean pCO_2 variability and dynamics. The clustering step ensures that clusters reflect areas of relatively coherent physical and/or biogeochemical behaviour much better than using the entire region as a whole (Landschützer et al., 2014; Rödenbeck et al., 2015; Gregor et al., 2019).

Clustering choice 1: Fay and McKinley (2014) biomes

The first empirical method for clustering the Southern Ocean is based on the global ocean regions or biomes (Figure 4.3) as defined by Fay and McKinley (2014) and extended by Gregor et al. (2019). Fay and McKinley (2014) created 17 open-ocean biomes using climatologies from four observational datasets: maximum mixed layer depth, sea-ice extent, chlorophyll-a and sea surface temperature. Specifically, with the inter-annual variability of these four inputs, they created dynamic biomes boundaries that move annually and a core biome map whose grid cells do not change biome assignment over the 13 years of the time-varying biomes. Based on their geographical locations and extent, Gregor et al. (2019) assigned manually the unclassified regions from the original biomes. Thus, this resulted in six additional regions as shown in Figure 4.3. These additional regions were not classified in the original biomes because the set of thresholds from Fay and McKinley (2014) did not consider the physical and biogeochemical properties. However, in this study, I use regionally fixed biome boundaries in a manual attribution, which is somehow similar to supervised clustering (Vichi et al., 2011). The resulting regionally fixed biomes are defined for the study region of interest which is the Southern Ocean south of 40°S.

Clustering choice 2: seasonal cycle reproducibility (SCR)

The second empirical approach used to divide the Southern Ocean into clusters hinges on analysing similar variability of surface ocean pCO_2, both zonally and seasonally. More specifically, this approach leverages mainly the SCR of the surface ocean pCO_2. The SCR is defined as the correlation of a detrended time series with its own climatology (Thomalla et al., 2011; Gregor et al., 2019). Using 10 years of the 5-day

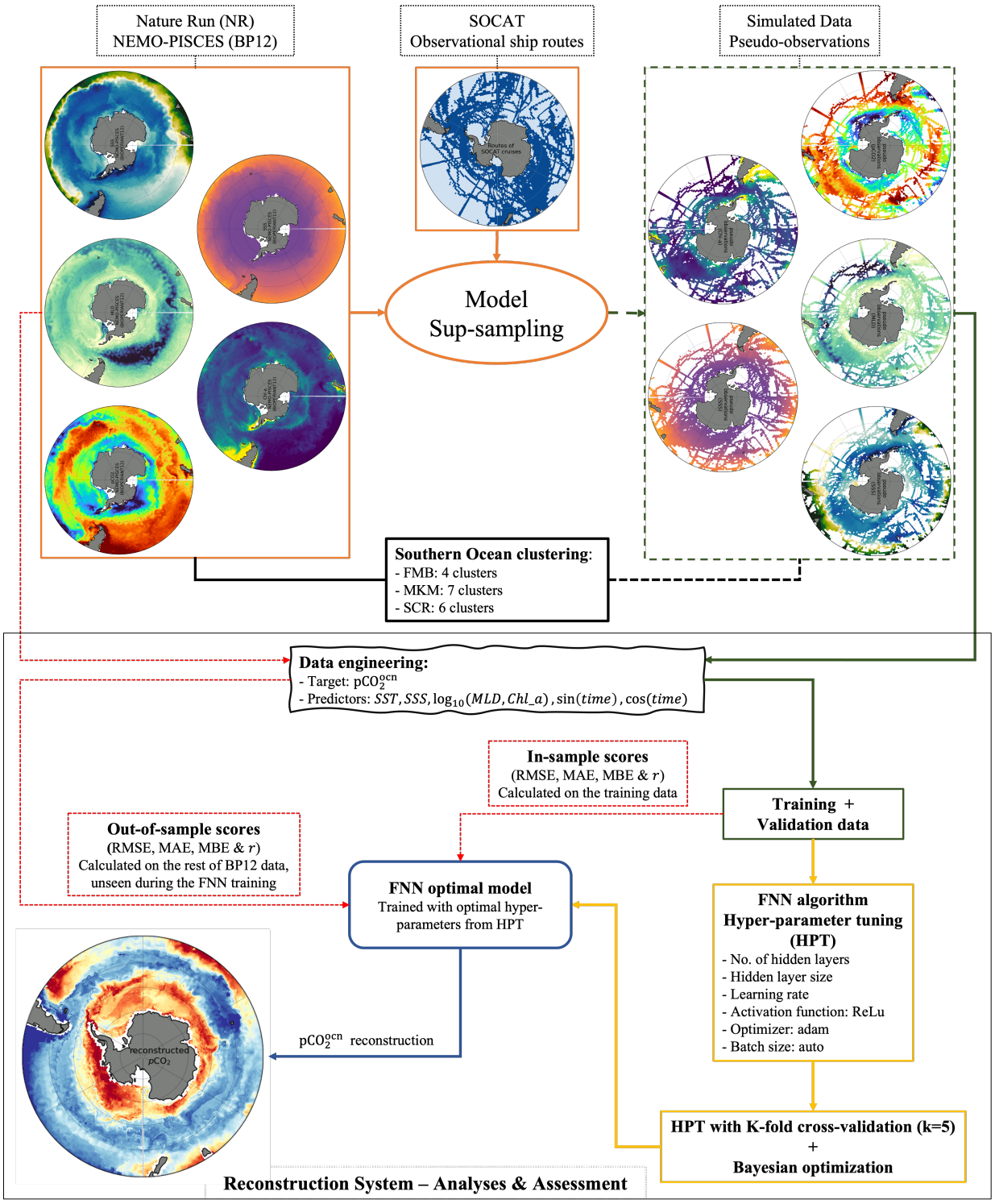


Figure 4.2: Schematic flow diagram of the implementation of the OSSE framework used in this study. It also shows various stages which are required in the three two-step approaches to reconstruct the surface ocean pCO_2 in the Southern Ocean. The implementation was done using the `Python` programming language in addition to some of its key packages such as `scikit-learn` and `scikit-optimize`.

ocean model output, the SCR of the surface ocean pCO_2 in the Southern Ocean south of 40°S is depicted in Figure 4.4. To account for a bit of the physical processes controlling the surface mixed layer pCO_2 due to the Antarctic Circumpolar Current (which is arguably the mightiest current in the oceans), the SCR map is separated into two zones based on the polar front (PF although the SCR is calculated grid point by grid

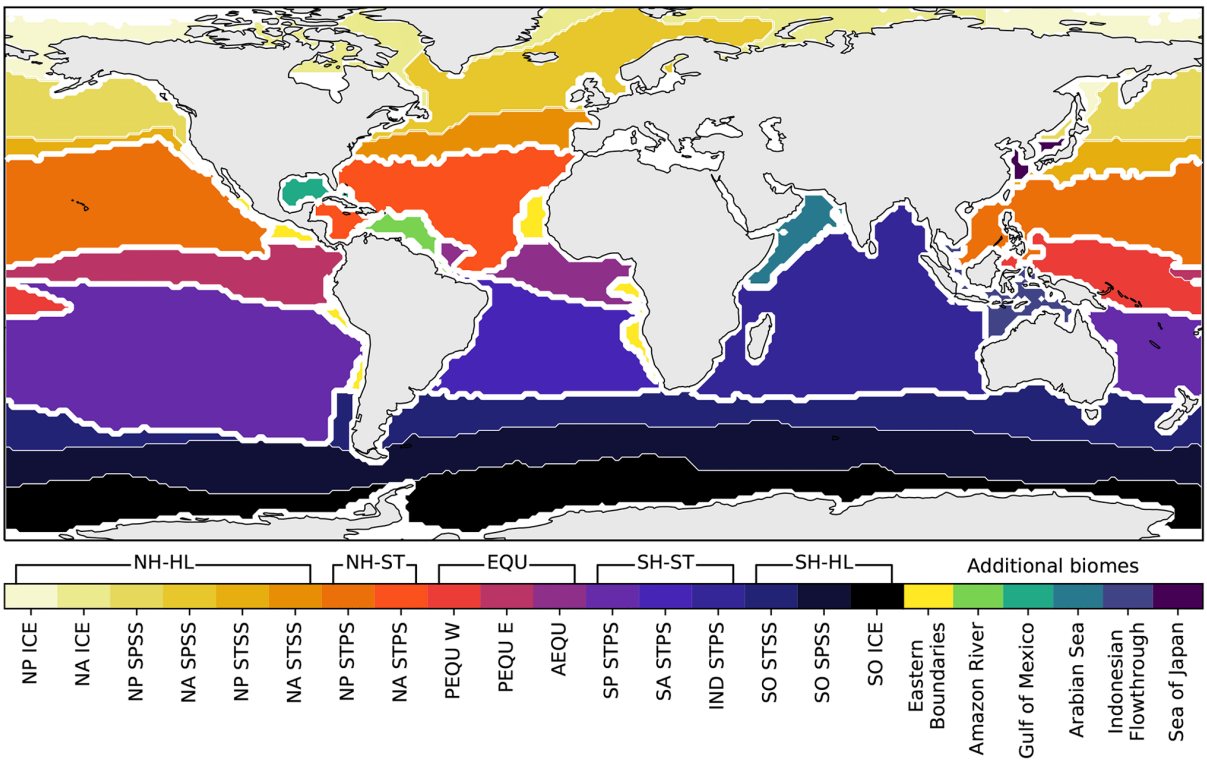


Figure 4.3: The global ocean biomes or regions as defined by Fay and McKinley (2014) and extended by Gregor et al. (2019). The thick white lines show the boundaries of the grouped regions (above the colour bar) where: NH stands for the Northern Hemisphere, SH stands for the Southern Hemisphere, HL is for high latitudes, ST is subtropics, and EQU stands for equatorial. See Fay and McKinley (2014) for the biome abbreviations below the colour bar.

point. Starting from the southern pole (Figure 4.4), there is the polar zone (PZ) which is the region south of the PF (Moore et al., 1999; Lovenduski et al., 2008), and the subpolar zone (SPZ) which is defined here as the region north of the PF.

The SCR of surface ocean pCO_2 (Figure 4.4) is equivalent to the percentage of the variance explained (commonly called R-squared or R^2) by the mean seasonal cycle of pCO_2, which explained how well the climatological mean seasonal cycle from the 10 years represents the evolution of surface ocean pCO_2 over each year (Figure 4.5). In areas of high SCR ($R^2 \geq 0.85$), there is neither intra-seasonal variability nor annual signals. In areas of medium SCR, the intra-seasonal variability starts emerging, however, in comparison to the seasonal cycle this variability is smaller in magnitude, while in areas of low SCR the seasonal signal is absent, and the intra-seasonal variability is larger than the seasonal cycle. Based on these criteria, the SCR of surface ocean pCO_2 is divided into three regions including (1) low-SCR (LSCR) that corresponds to the range $[0.0, 0.65)$; (2) medium-SCR (MSCR) corresponding to $[0.65, 0.85)$; and (3) high-SCR (HSCR) that corresponds to SCR values belonging to the interval $[0.85, 1.0]$.

Clustering choice 3: Mini batch K-Means

The mini-batch K-Means (MKM) algorithm is the third clustering approach used to separate the Southern Ocean into clusters. Derived from the popular machine learning algorithm K-Means, MKM is one of the

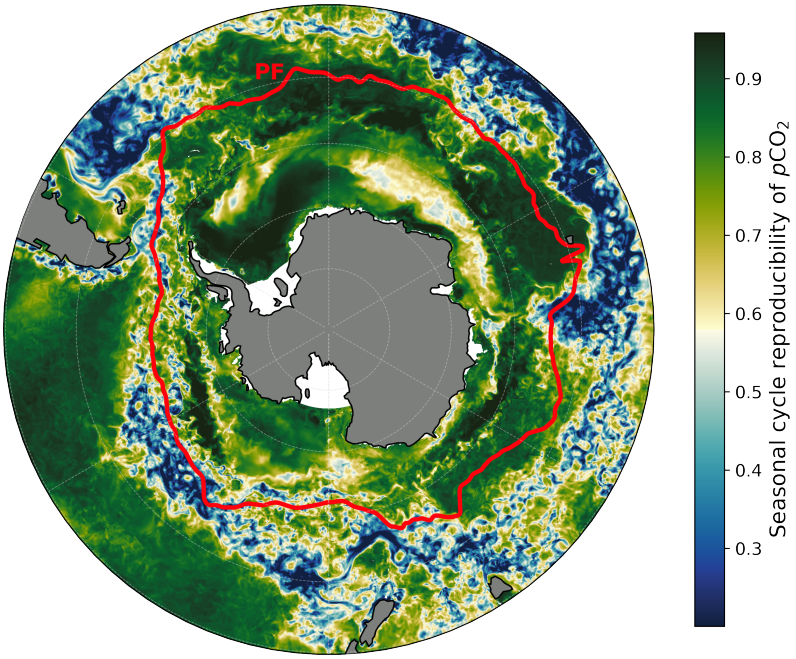


Figure 4.4: The seasonal cycle reproducibility (SCR) of pCO_2 calculated using NEMO-PISCES 5-daily outputs of 10 years. This is a much lower temporal resolution (5 days) compared to the one day of BIOPERIANT12 (available for only one year) used throughout this work. The SCR is defined as the correlation of the detrended pCO_2 with its own climatology, whose from definition the calculation requires more than one year data; the larger the correlation, the stronger the reproducibility of the seasonal cycle (Thomalla et al., 2011). The polar front (PF) is depicted by the red line.

widely used unsupervised machine learning algorithms for clustering and pattern recognition. In Figure 4.6, the graphics compare the execution time of K-Means (Figure 4.6a) and mini-batch K-Means (Figure 4.6b) approaches. For many decades, K-Means has been one of the most used clustering algorithms, mainly because of its good time performance (Hartigan and Wong, 1979; Jain, 2010). However, with the recent increase in sizes and volumes of datasets subjected to analysis, the K-Means approach has been losing its attractiveness because of its constraint of requiring the entire dataset in memory during the training phase of the model.

In this MKM clustering method, clusters are created from the configuration of the surface ocean pCO_2 and its proxy variables including the sea surface temperature, sea surface salinity, mixed layer depth, and chlorophyll-*a* concentration, all from the training dataset or pseudo-observations. Besides using a performance metric to select a configuration, an optimal cluster configuration must contain the pCO_2 variable and a combination of its proxy variables to ensure that the underlying relationship existing between the surface ocean pCO_2 and its proxies (Takahashi et al., 1993) is preserved for the regression step.

Second step: regression modelling of pCO_2

In this two-step approach, the second step consists of using a machine learning regression method inside clusters resulting from the first step of the approach. The named machine learning (ML) model is the feed-

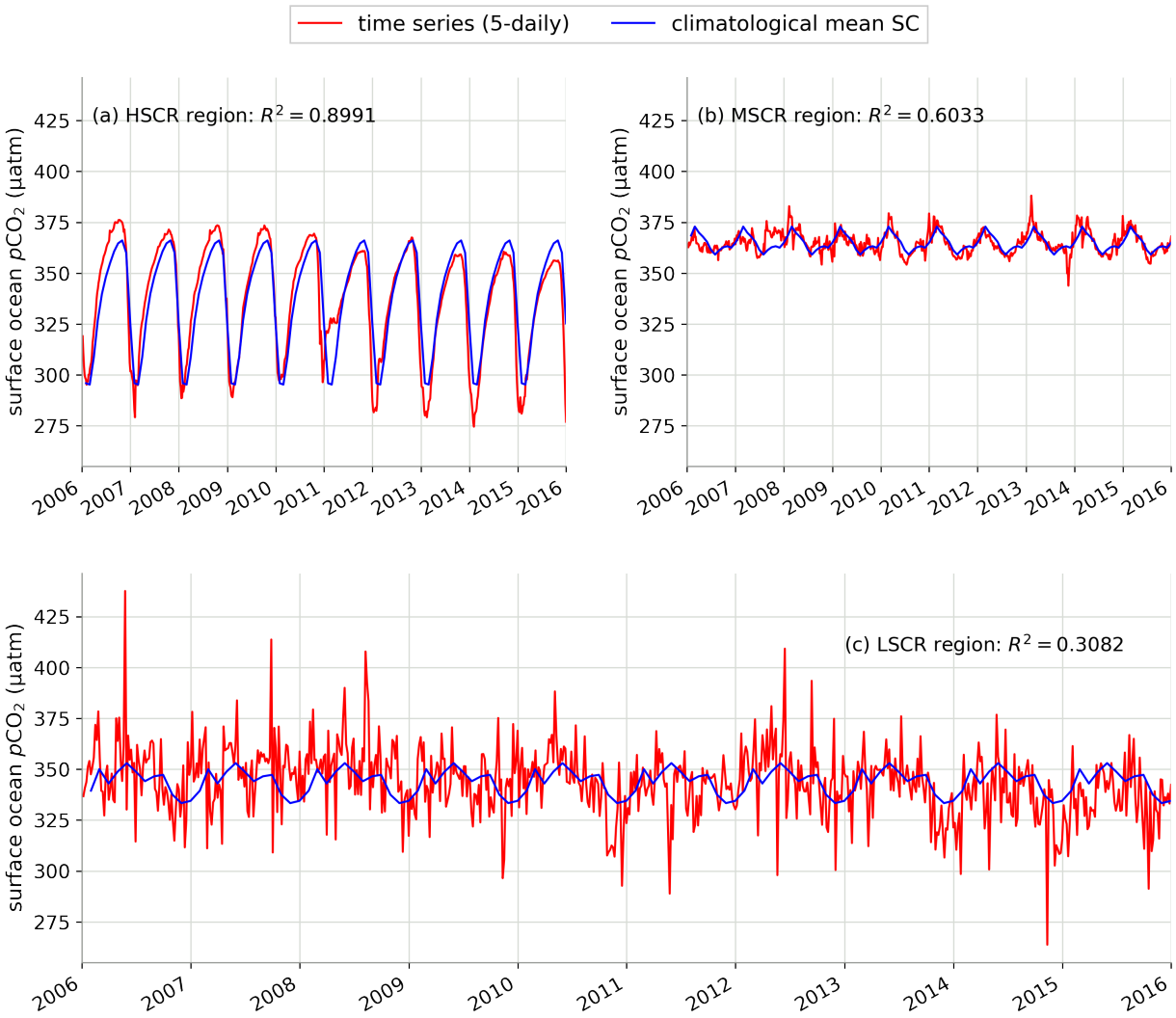


Figure 4.5: Time series of surface ocean pCO_2 from 2006 to 2015 (in red) compared to its climatological mean seasonal cycle (in blue) calculated over the 10 yr for three $1/12°$-by-$1/12°$ grid cells corresponding to HSCR or panel **(a)** a region of high SCR in relatively low pCO_2 subpolar zone (SPZ), MSCR or panel **(b)** a region of medium SCR in relatively low pCO_2 SPZ, and panel **(c)** a region of low SCR in high pCO_2 polar zone (PZ). With all correlations being statistically significant at a 95% confidence level, the percentage of variance explained by the regression of the time series (in red) onto the climatological mean seasonal cycle (SC, in blue) is shown as R^2 in the figure.

forward neural network (FNN) whose algorithm is presented in Section 2.3.2. Details on the regression modelling of surface ocean pCO_2 are given in Section 2.4.

Usually, for the ML model development including training and validation, the data are randomly divided with an appropriate ratio (e.g., 0.8 : 0.2) into a training dataset and a test dataset. However, in this study, it is not necessary to do so because of the advantage of using observing system simulation experiments (OSSEs). This advantage is that in the OSSE framework, we have access to the whole study domain data, which therefore allows a thorough assessment of the model performance on data that have not been involved in creating the synthetic data used to train the ML model.

The major challenge in ML model development is to overfit the model and then over-interpolate the data. To further reduce the possibility of overfitting, hyper-parameters of the FNN algorithm are fine-tuned for

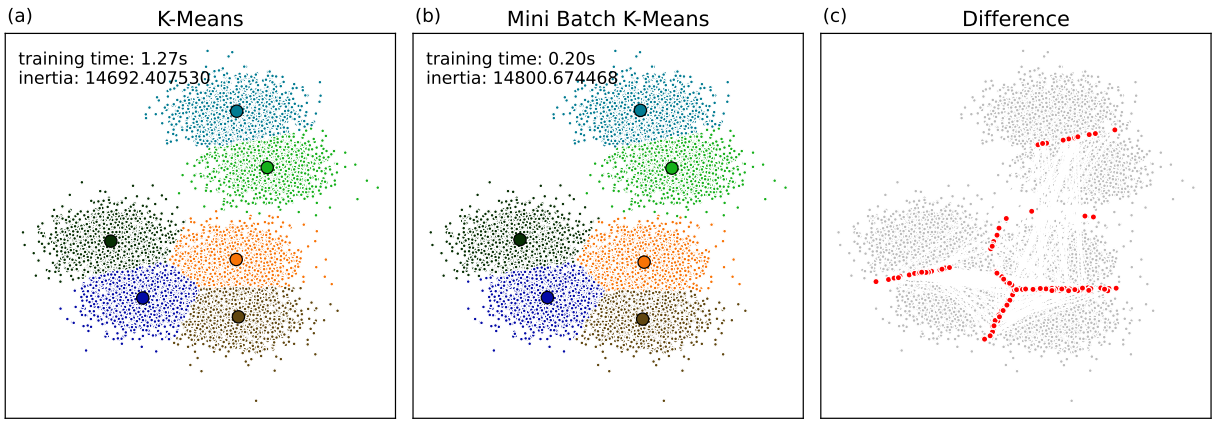


Figure 4.6: K-Means vs Mini Batch K-Means: a schematic illustration of the performance of the two approaches in terms of their training-phase times. It shows that Mini Batch K-Means clustering outperformed the K-Means clustering algorithm. Recent studies of the performance of these two methods came to a similar conclusion (Béjar, 2013; Feizollah et al., 2014; Hicks et al., 2021).

the model to be relatively generalisable; that is, able to interpolate well the data that the model has not been exposed to during training. Therefore, the hyper-parameter optimization is achieved using a combined method that consists of Bayesian optimization and K-fold cross-validation for which a subset of the training data is iteratively left out from the training process for a certain set of hyper-parameters (detail in Section 2.3.7). The hyper-parameters that result in the best score from this optimization process are then used for the final model fitting with the full training data.

4.3 Results and analysis

4.3.1 Pseudo-observational results

Underway SOCAT ship observations of surface ocean pCO_2 and proxies, called pseudo-observations, were sub-sampled from the Nature Run of the OSSE framework (Section 4.2.3) using the underway SOCAT ship routes described in Section 4.2.2. As results from this SOCAT data sampling simulation in the Southern Ocean, Figure 4.7 shows the annual mean of these pseudo-observations: surface ocean pCO_2 (Figure 4.7a) and its main proxy variables (Figure 4.7b-e).

4.3.2 Southern Ocean clustering results

Figure 4.8 shows the results of the clustering of the Southern Ocean south of 40°S as described in Section 4.2.4.1.

FMB-based clusters

Table 4.1 and Figure 4.8a present the four resulting sub-regions of the study area (i.e., the Southern Ocean

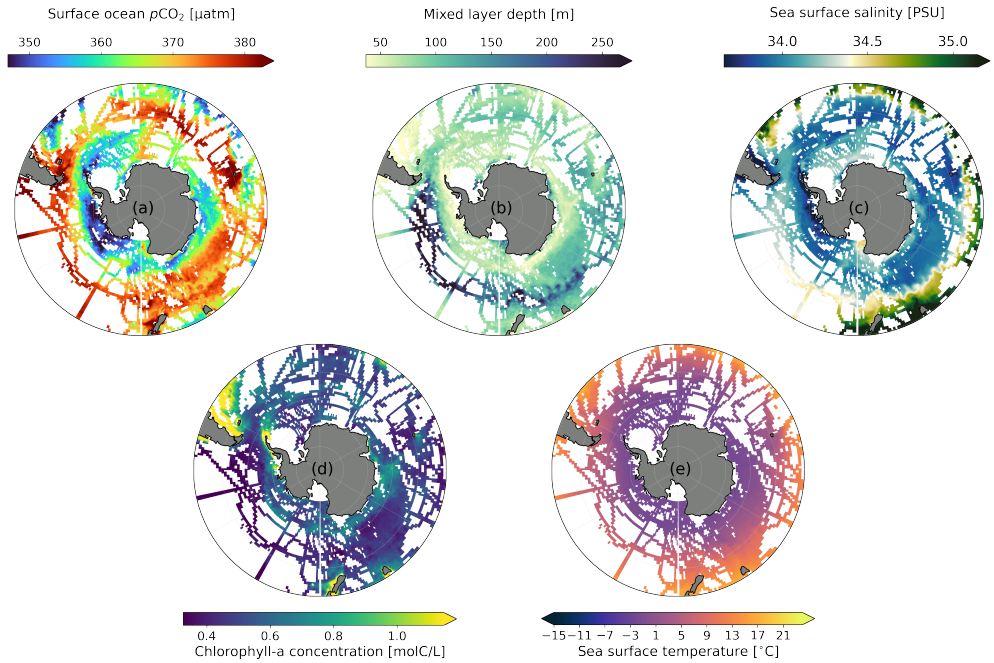


Figure 4.7: Maps of mean values resulting from using the OSSE Nature Run to back out pseudo-observations from SOCAT ship routes in the Southern Ocean (south of 40°S): **(a)** surface ocean pCO_2, **(b)** mixed layer depth (MLD), **(c)** sea surface salinity (SSS), **(d)** chlorophyll-a concentration (Chl-*a*), and **(e)** sea surface temperature (SST).

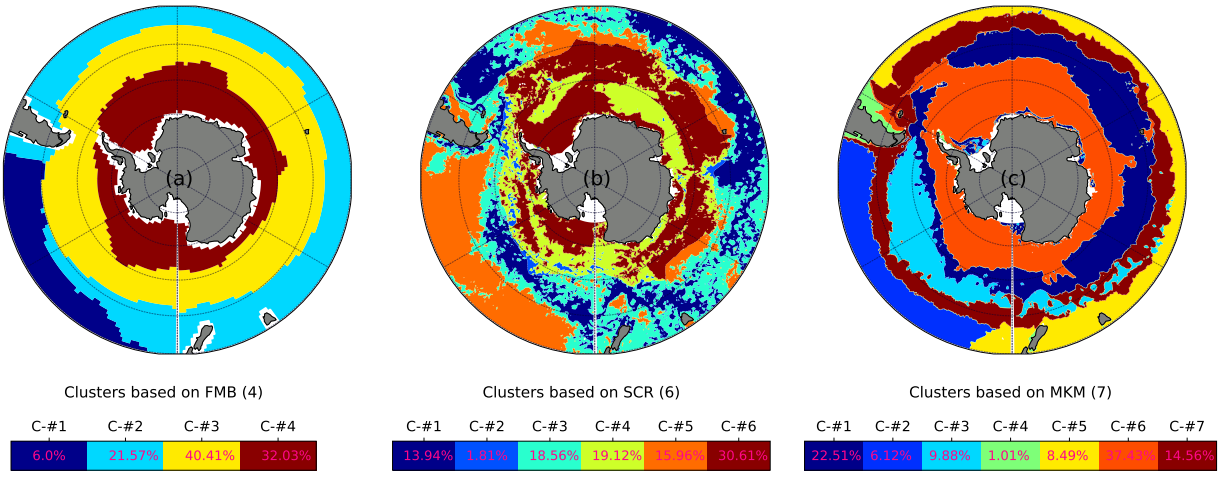


Figure 4.8: Maps of the Southern Ocean (south of 40°S) sub-regions. **(a)** Fay and McKinley (2014)'s biomes (FMB, see Figure 4.3), panel **(b)** seasonal cycle reproducibility (SCR) (see Figure 4.4), and **(c)** mini-batch K-Means (MKM) clustering method. The percentage (%) in the colour bar indicates the fractional coverage for each sub-region relative to the whole region. For the sake of readability, the named sub-regions from the three methods are presented with sequential numbers (C-#.), and described in Tables 4.1 and 4.2 for FMB and SCR, respectively.

south of 40°S) and the estimates of the fractional coverage (%) of each of these sub-regions relative to the whole region (colour bar in Figure 4.8a).

The largest FMB-based cluster is the SO SPSS biome with about 40.41%coverage relative to the whole study area. The common names for these four biomes are equivalent to the sub-Antarctic zone (SAZ) combining the SP STPS and SO STSS biomes, the polar frontal zone (PFZ) for the SO SPSS biome, and the seasonally ice-covered zone for the SO ICE (Lovenduski et al., 2008).

Table 4.1: Characteristics of Fay and McKinley (2014)' biomes that were used for clustering the Southern Ocean south of 40°S.

Cluster No.	Abbreviation	Description	Fraction coverage (%)
C-#1	SO ICE	Southern Ocean Ice	32.02
C-#2	SO SPSS	Southern Ocean Subpolar Seasonally Stratified	40.41
C-#3	SO STSS	Southern Ocean Subtropical Seasonally Stratified	21.57
C-#4	SO STPS	Southern Ocean Subtropical Permanently Stratified	6.0

SCR-based clusters

Using the 2-by-3 level clustering approach based on the seasonal cycle reproducibility (SCR) of the surface ocean pCO_2 (described in Section 4.2.4), the Southern Ocean south of 40°S was also divided into six distinct sub-regions whose characteristics are presented in Table 4.2 and Figure 4.8b. Results show the HSCR was the largest sub-region in the polar zone (PZ) and the MSCR was the largest sub-region in the subpolar zone (SPZ) with respectively about 30.65% and 19.56% coverage relative to the whole Southern Ocean.

Table 4.2: Description of the Southern Ocean (40°S) clusters from the seasonal cycle reproducibility (SCR) of the surface ocean pCO_2 and the polar front (PF) (Figure 4.3).

Cluster No.	Abbreviation	Description	Fraction coverage (%)	SCR range
C-#1	LSCR SPZ	Low SCR Subpolar Zone	13.94	[0.0, 0.65)
C-#2	LSCR PZ	Low SCR Polar Zone	1.81	
C-#3	MSCR SPZ	Medium SCR Subpolar Zone	18.56	[0.65, 0.85)
C-#4	MSCR PZ	Medium SCR Polar Zone	19.12	
C-#5	HSCR SPZ	High SCR Subpolar Zone	15.96	[0.85, 1.0]
C-#6	HSCR PZ	High SCR Polar Zone	30.65	

The percentage of the variance explained by the seasonal cycle of surface ocean pCO_2 defines how well the mean climatological seasonal cycle (e.g., from one year) represents the evolution of surface ocean pCO_2 over each year. Areas where the seasonal cycle for each year is coherent with the year series (e.g., Figure 4.5a), are defined as having high seasonal cycle reproducibility (SCR ≥ 0.85). Regions where there is large variability from year to year in the timing and amplitude of the surface ocean pCO_2 (e.g., Figure 4.5c), and only a low percentage of the variance can be explained by the mean seasonal cycle are defined as having low SCR (SCR < 0.4). Thus, immediately apparent from Figure 4.4 are the sharp gradients between strongly contrasting regions of high ($> 70\%$) and low ($< 30\%$) seasonal cycle reproducibility.

MKM-based clusters

Figure 4.9 shows various configurations that were used in the Mini-batch K-Means (MKM) algorithm, the third clustering method used to create the Southern Ocean clusters. These cluster configurations were assessed using both the Sum of Squared Error (SSE) also known as Inertia (Figure 4.9a) and the Silhouette score (Figure 4.9b). Both the Silhouette score and the Inertia are assessment metrics used to measure the goodness of a clustering technique. Overall, the Silhouette score of a cluster is based on the comparison of its tightness and separation, and shows which objects lie well within the cluster, and which ones are merely somewhere in between clusters (Rousseeuw, 1987).

The Silhouette score values fall within the range $[-1, 1]$, where for a chosen configuration, a score of 1 means that the clusters are very dense and well separated; a score of 0 means that the clusters are overlapping; while a negative score means that something is not right with the configuration used or the data involved. With this criterion, the best clustering configuration including the data variables used (see legends, Figure 4.9) was selected and subsequently, applying the Elbow Criterion to the Inertia graph, an optimal number of clusters was chosen. This criterion consists of choosing a relatively small number of clusters (K) that still has lower Inertia, which usually represents where diminishing returns start to happen with increasing K.

As shown in Figure 4.9a, for all the cluster configurations, the Inertia decreased abruptly until $K = 7$ from which the slopes get much higher. Combining with the Silhouette score (Figure 4.9b), the optimal setting for $K = 7$ corresponded to the cluster configuration “abcde”, which resulted from the MKM clustering of the Southern Ocean using the following data variables (cf. keys in Figure 4.9b): surface ocean pCO_2, sea surface temperature (SST), sea surface salinity (SSS), mixed layer depth (MLD), and chlorophyll-*a* (Chl-*a*). This is a mixed subjective choice because the Silhouette score shows flattening beyond $K = 9$, however, for $K = 9$ this would correspond to clusters created without any knowledge of the SSS and Chl-*a* as this would correspond to the cluster configuration "abd" while $K = 8$ would not be an option either as this would mean clusters without knowledge of the surface ocean pCO_2.

Figure 4.8c shows the map of Southern Ocean (40°S) clusters which resulted from using the optimal configuration "abcde" + (K = 7). These seven resulting clusters were labelled as follows: C-#1, C-#2, C-#3, C-#4, C-#5, C-#6, and C-#7, where estimates of their fraction coverage (Figure 4.8c) relative to the whole Southern Ocean south of 40°S are indicated in percentage (%). Each of these clusters reflects an area where some underlying physical and/or biogeochemical relationship exists between the surface ocean pCO_2, SST, SSS, MLD, and Chl-*a*. The two largest clusters are C-#6 and C-#1 with the fraction coverages estimated at 37.43% and 22.51%, respectively; while the two smallest clusters are C-#4 and C-#2, which respectively cover 1.01% and 6.12% of the whole study domain (Figure 4.8c).

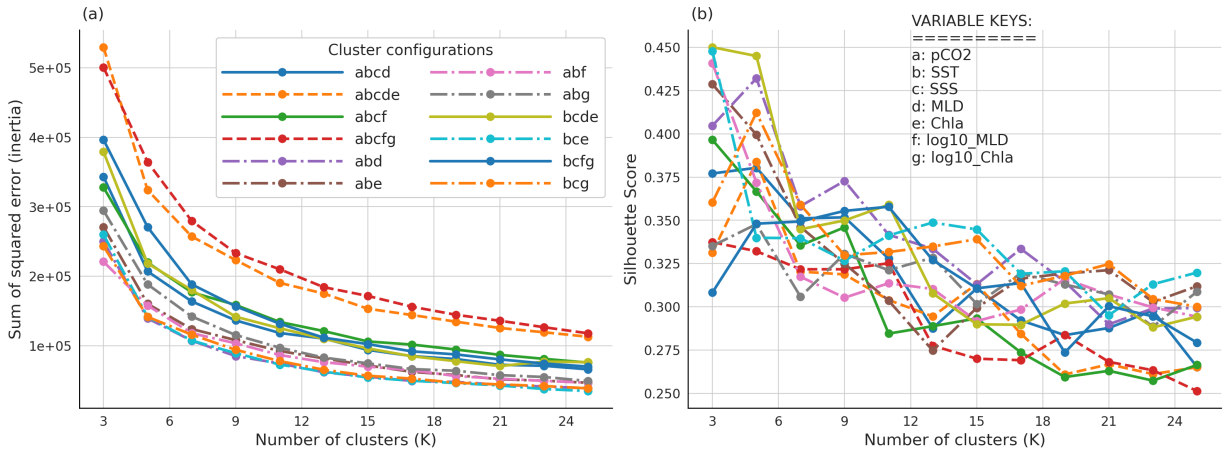


Figure 4.9: Mini-batch K-Means (MKM) clustering evaluation and selection of the optimal clusters. Here, **(a)** is the plot of the Sum of Squared Error (SSE) or the sum of the squared distance between each member of the cluster and its centroid against the number of clusters (K), and **(b)** is the plot of the Silhouette scores. The legend in **(a)** must be read together with that of **(b)**.

4.3.3 Performance analysis

The main advantage of coupling observing system simulation experiments (OSSEs) and machine learning is that both in- and out-of-sample scores can be estimated. The in-sample score is calculated from the synthetic training data points, which result from the SOCAT-ship pseudo-observations; while the out-of-sample error is calculated from the entire predicted domain data (Gregor et al., 2017). The out-of-sample score gives a better representation of the true error of the method. Both types of scores serve to provide more insights regarding the goodness of fit or prediction of the underlying empirical regression method. Based on the underway SOCAT ship observations (2020 release) in the Southern Ocean south of 40°S, the sampled grid cells represent only 60% of the whole region. Data corresponding to this 60% are thus used for in-sampling scores or goodness of fit (an indicator of the performance of regression models on training data) while data from the remaining 40% grid cells, where ships have not yet reached, are used to calculate the out-of-sample scores. The root mean squared error (RMSE) is considered here as the main metric for analysis of methods' performance. Results from various OSSE runs are presented in the following sections.

Model validation diagnostics: error analysis

The training errors are diagnosed here to better understand and assess how well, during the development of the underlying machine learning model, each mapping approach did across the two OSSE scenarios. The goodness of fit was then evaluated by the training of FMB, MKM and SCR learning models with the pseudo-observations, which resulted from backing out synthetic SOCAT data from the Nature Run field. For each of the two OSSE scenarios including SOCAT-only and SOCAT + WGM (Section 4.2.2), I present here the results from the analysis of the method performance on the training dataset. In Figure 4.10, the top row illustrates the validation diagnostics for the machine learning models trained with synthetic data from

the SOCAT-only scenario, while the bottom row shows the diagnostics when these ML models were trained with synthetic data from the SOCAT + WGM scenario.

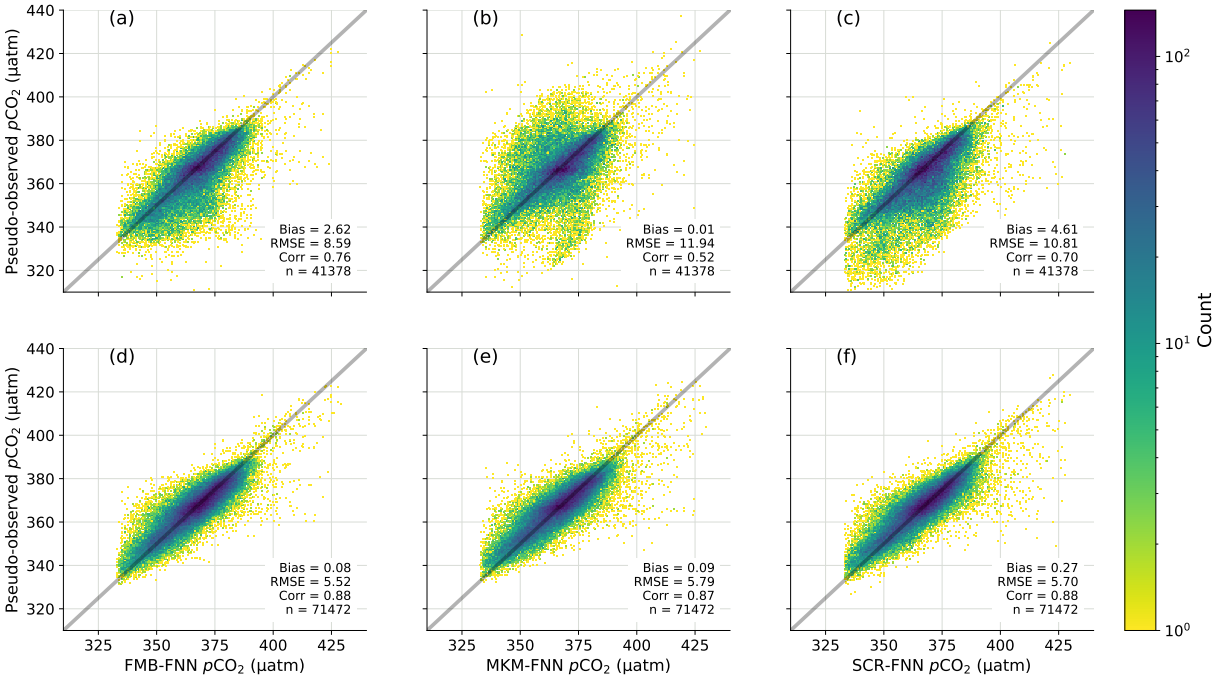


Figure 4.10: Diagnostic training plots (predicted vs. pseudo-observed) for the two experimental regimes. pCO_2. The top row corresponds to the SOCAT-only regime and the bottom row is the SOCAT + Winter Gap Mitigation (WGM) regime. The first, second, and third columns respectively show the diagnostics of FMB-FNN, MKM-FNN, and SCR-FNN methods trained with synthetic from the two regimes.

Figures 4.10a-c show the scatter plots of the surface ocean pCO_2 estimates against the ground truth for all three gap-filling approaches (Section 4.2.4). In these figures, the grey line represents what the perfect regression line would look like so that the closer the points are to this line, the more accurate the method is. The FMB learning method yielded the best correlation ($r = 0.75$), the SCR fell behind ($r = 0.70$), and the MKM performed the worst ($r = 0.52$) despite having the lowest bias (0.01 µatm). Compared to the variation of pCO_2 pseudo-observations, all the biases are relatively small, and their uncertainties characterized here by the root mean squared errors (RMSEs) are on the same order of magnitude for all three approaches. For the SOCAT + WGM scenario (Section 4.2.2), Figures 4.10d-f show the training diagnostics for all three gap-filling methods. It can be seen that the three methods performed almost equally on the training dataset with a minor difference in the estimates of their RMSE scores (Figures 4.10d-f). Considering the bias estimates, it can be remarked that it is instead slightly higher in the SCR learning method.

Out-of-sample scores

In more than 90% of supervised machine learning applications, models are intended to make predictions or reconstructions as in the case of this study. An optimal model resulting from training and validation is expected to perform "well" not only on the training dataset but also on unseen data in order to be generalizable. As mentioned early, in the context of OSSEs, the key advantage was that out-of-sample scores could

be calculated as the whole domain model data were accessible through Nature Run (Section 4.2.3). These out-of-sample scores allowed us to confidently evaluate and compare as follows the clustering choice from the perspective of the two OSSE scenarios and are summarized in Table 4.3. More specifically, this table shows the Southern Ocean and regional breakdown of out-of-sample scores measuring the performance of reproduction of surface ocean pCO_2 in both SOCAT-only SOCAT + WGM scenarios using the OSSE system as presented by the flow diagram (Figure 4.6).

Table 4.3: A summary of various out-of-sample scores or performance metrics (MBE or Bias, MAE, RMSE, and r). The RMSE scores for empirical estimates of pCO_2^{ocn} are calculated for the Southern Ocean (SO) and then all the zonal regions (described in Section 2.5 and defined by Lovenduski et al. (2008)) while the bias or mean bias error and Pearson's correlation are computed for the whole SO, using the various clustering approaches. Values in the table are significantly different from the mean for the column and clustering method, with a p-value (p)< 0.05 for the two-tailed Z test.

OSSE sampling scenaios	Mapping approach	Root mean squared error (µatm)				Bias (µatm)	MAE (µatm)	r
		SO	SAZ	PFZ	AZ			
SOCAT-only	FBM-FNN	9.93	13.12	6.50	7.90	3.53	6.69	0.69
	MKM-FNN	12.01	12.82	10.49	11.93	1.22	8.3	0.54
	SCR-FNN	10.85	10.10	11.44	11.27	4.74	7.66	0.70
SOCAT + WGM	FBM-FNN	5.97	6.71	5.42	5.31	0.40	4.29	0.84
	MKM-FNN	5.98	6.21	5.36	5.61	0.13	4.31	0.87
	SCR-FNN	5.83	6.03	5.16	5.52	0.09	4.18	0.88

The SOCAT-only scenario

Let us recall that in the SOCAT-only regime or scenario, synthetic SOCAT observations are biased towards summer like in their real-world counterpart (Bakker et al., 2016). Focusing on the mapping error RMSE, Table 4.3 shows that FMB-FNN performs better in the SO than the other two mapping methods. The SCR-FNN mapping approach yields the best Pearson's correlation (0.70) in the Southern Ocean (Table 4.3) by a small margin when considering the two other approaches.

Now, let us look at these mapping errors in more detail across regions. In the SAZ, SCR-FNN achieves an RMSE of 10.1 µatm, thus, outperforming MKM-FNN and FMB-FNN which individually achieves higher RMSE scores of 12.82 µatm and 13.12 µatm, respectively. This is also reflected in the mean bias error (MBE) estimates (Table 4.3). However, in the PFZ and AZ, the analysis gives the same conclusion as for the Southern Ocean as a whole. In these two regions, the FMB-FNN approach outperforms the MKM-FNN and SCR-FNN by relatively large margins (when considering the RMSE score) estimated respectively at 4.1 µatm and 4.94 µatm in the PFZ, and respectively at 4.03 µatm and 3.37 µatm in the AZ.

The top rows of Figures 4.11 and 4.12 present the maps of the reconstruction biases and RMSEs, respectively. It can be seen in Figure 4.11a-c that the spatial distribution of the bias is very heterogeneous throughout the Southern Ocean as a whole regardless of the methods. Similar remarks are made regarding

the spatial distribution of RMSEs in Figure 4.12a-c. This heterogeneity is likely due to the clustering step which may be acting differently on the matching between pCO_2 and proxies across various clusters. Regionally, the annual mean of surface ocean pCO_2 from FMB-FNN, MKM-FNN, and SCR-FNN can be biased high or low by more than 10 µatm (Figure 4.11a-c). However, these patches average out in some parts of the sub-Antarctic zone (SAZ), polar frontal zone (PFZ) and Antarctic zone (AZ) in such a way that the mean bias error becomes smaller than 5 µatm. Biases are larger in the SAZ, especially in the Pacific Ocean sector where ship-based observations are less dense, and in some local areas of the AZ as shown in Figure 4.11a-c.

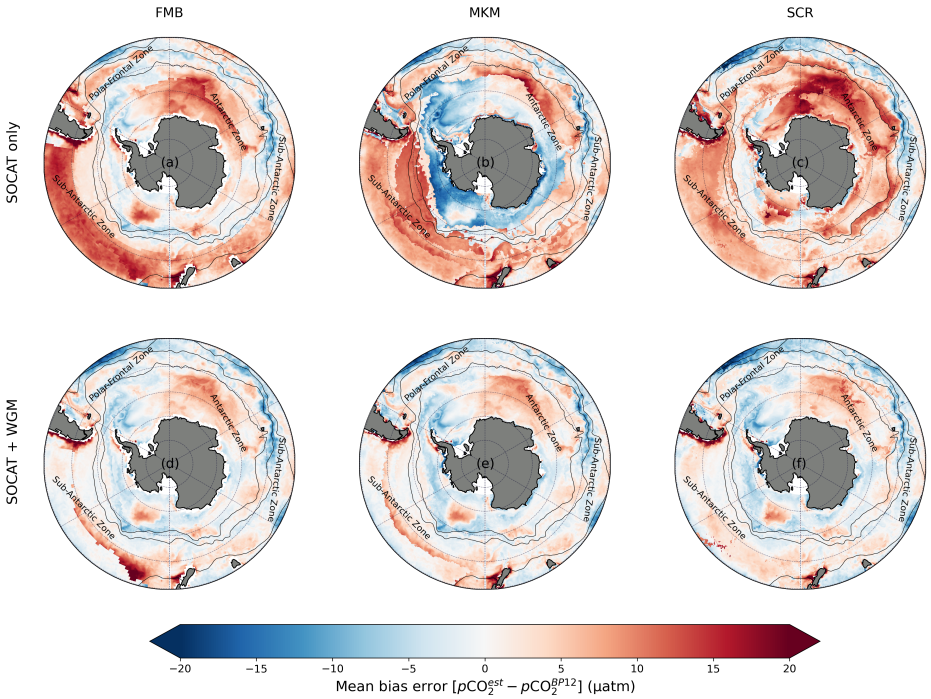


Figure 4.11: Maps of mean bias errors (MBEs) of pCO_2 in both SOCAT-only and SOCAT + WGM experiments. The top row shows the MBE maps of FMB-FNN, MKM-FNN and SCR-FNN trained with synthetic SOCAT data only while the bottom row corresponds to the MBE maps of these three methods trained with synthetic SOCAT data plus pseudo-observations from one winter cruise on SOCAT-ship lines.

Figure 4.12a-c shows regionally significant regional variability of the RMSE scores, which can be larger, especially in some poorly sampled regions such as the Pacific Ocean sector of the SAZ. In this SOCAT-only scenario, however, this RMSE regional variability is also method-dependent. For example, the FMB-FNN and MKM-FNN approaches seem to have larger RMSEs in the Pacific Ocean sector than any other areas of the Southern Ocean (Figure 4.12a-b) while the SCR-FNN approach shows larger RMSEs in the Atlantic Ocean sector (Figure 4.12c).

The SOCAT + WGM scenario

Figures 4.11d-f and 4.12d-f show the spatial distribution of the mean bias errors (MBEs) and the root mean squared errors (RMSEs) of these mapping methods, while the overall regional performances are summarized in Table 4.3. By supplementing the synthetic SOCAT-only observations with pseudo-winter (SOCAT +

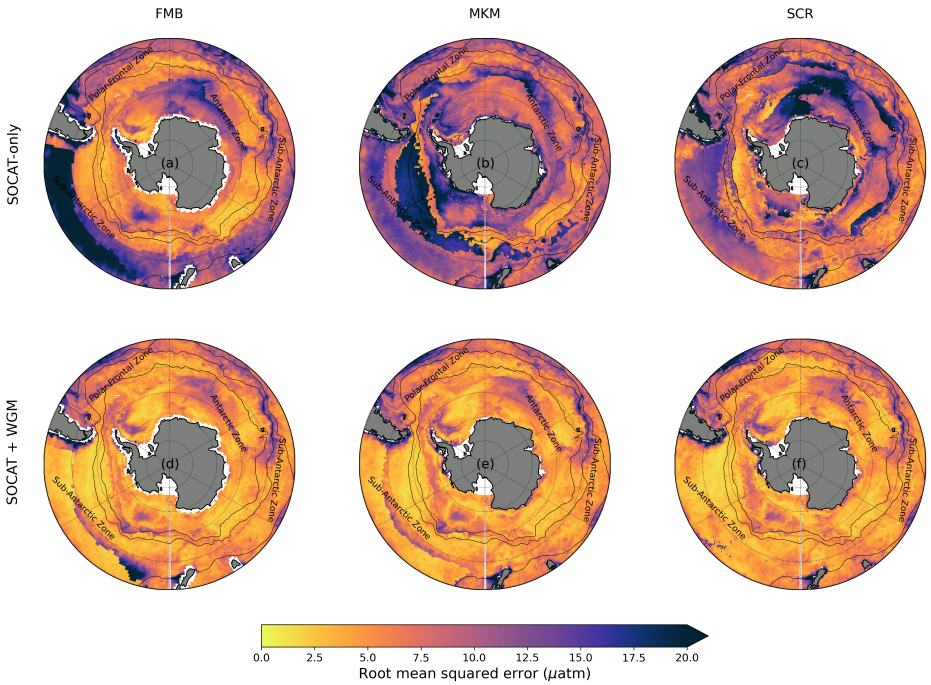


Figure 4.12: Uncertainty maps in terms of root mean squared errors (RMSEs) of pCO_2 reconstructions in both SOCAT-only and SOCAT + WGM experiments. The top row shows the MBE maps of FMB-FNN, MKM-FNN and SCR-FNN trained with synthetic SOCAT data only while the bottom row corresponds to the MBE maps of these three methods trained with synthetic SOCAT data plus pseudo-observations from synthetic winter measurements on SOCAT-ship lines.

WGM, Section 4.2.2) in the Southern Ocean, we explored the potential impact that the winter gap in SOCAT data has on surface ocean pCO_2 estimates in the region. Except for the FNN-based methods (Table 4.3), the out-of-sample scores indicate significant improvements across all the Southern Ocean regions, especially in the sub-Antarctic zone (SAZ) of the Pacific Ocean sector, which recorded much lower RMSE scores (Figure 4.12d-f) than in the SOCAT-only scenario (Figure 4.12a-c).

Overall, for both OSSE regimes SOCAT-only and SOCAT + WGM, the performance of each of the three mapping methods depends on Southern Ocean regions (cf. Table 4.3). This suggests that each region contributes spatially differently to the performance of each method in the whole Southern Ocean. For example, the FMB learning method outperforms when considering the SOCAT-only regime while the SCR learning method outperforms when considering the SOCAT + WGM scenario.

4.3.4 Ensemble average results

Individually, in our previous analyses, the mapping methods demonstrated in each of the OSSE runs their strengths and weaknesses when reproducing the surface ocean pCO_2 in the Southern Ocean south of 40°S. Regional performance analysis led to slightly different conclusions across various clustering-based methods (cf. Table 4.3). While the FMB-FNN approach seemed to outperform MKM-FNN and SCR-FNN in the SOCAT-only scenario, it is the SCR-FNN approach that outstood in the SOCAT + WGM scenario as shown

in Table 4.3. Therefore, it becomes difficult to tell which one of the three approaches superior to evaluate the two OSSE scenarios. One possible option is to consider a three-member ensemble average technique which consists of taking the average of the pCO_2 estimates from the three clustering-based methods. In such a situation, the ensemble average estimates have been proven to be more effective than using estimates from individual methods (Gregor et al., 2019).

Table 4.4 shows the results of the ensemble average (ENS) estimates. These results show better performance metrics in both OSSE scenarios compared to the individual approaches (cf. Table 4.3). According to all the evaluation metrics RMSE, MAE, Bias and r, the reconstruction uncertainties and biases from the ENS show important improvements either in the SOCAT-only scenario where the observational gaps are the largest in winter or in the new observing scenario SOCAT + WGM where winter gaps are mitigated. In other words, either in the whole Southern Ocean or regionally, the pCO_2 reconstruction significantly improved in the SOCAT + WGM scenario compared to the SOCAT-only scenario by relatively large margin estimates of, for example, 3.19 µatm (RMSE), 3.08 µatm (Bias), 2.2 µatm (MAE), and $0.12(r)$, when considering the Southern Ocean (Table 4.4).

Table 4.4: Performance metrics (MBE or Bias, MAE, RMSE, and r) for the ensemble average (ENS) of pCO_2^{ocn} estimates resulting from the FMB-FNN, MKM-FNN, and SCR-FNN approaches for both SOCAT-only and SOCAT + WGM scenarios. Any other characteristics of this table are smilarly presented in Table 4.3's caption.

OSSE sampling scenaios	Region	Bias (µatm)	MAE (µatm)	RMSE (µatm)	r
SOCAT-only	SO	3.23	6.25	8.72	0.73
	SAZ	5.12	7.71	10.87	0.61
	PFZ	2.69	5.69	7.79	0.59
	AZ	2.33	5.35	7.12	0.79
SOCAT + WGM	SO	0.15	4.05	5.53	0.85
	SAZ	0.59	4.03	5.80	0.82
	PFZ	0.43	3.65	5.06	0.78
	AZ	0.26	3.97	5.15	0.86

The mean bias error (MBE) maps resulting from the ENS estimates (Figure 4.13a, c) show that there have been significant improvements in the SOCAT-only and SOCAT + WGM scenarios compared to the individual methods (Figures 4.11). For example, the ENS estimates significantly improve the MBEs in the Pacific Ocean sector of the SAZ (cf., Figure 4.11), while the RMSEs remain similar (cf. Figure 4.12). The MBE and RMSE combined with the correlation indicate the degree to which the matching of estimates and ground truth is consistent across Southern Ocean regions (Table 4.4; Figure 4.13). The spatial coincidence of low MBEs (Figure 4.13a, c) and low RMSEs (Figure 4.13b, d) combined with high correlations (Table 4.4) indicates that the pCO_2 reconstruction performs well across all conditions represented by this specific

Nature Run.

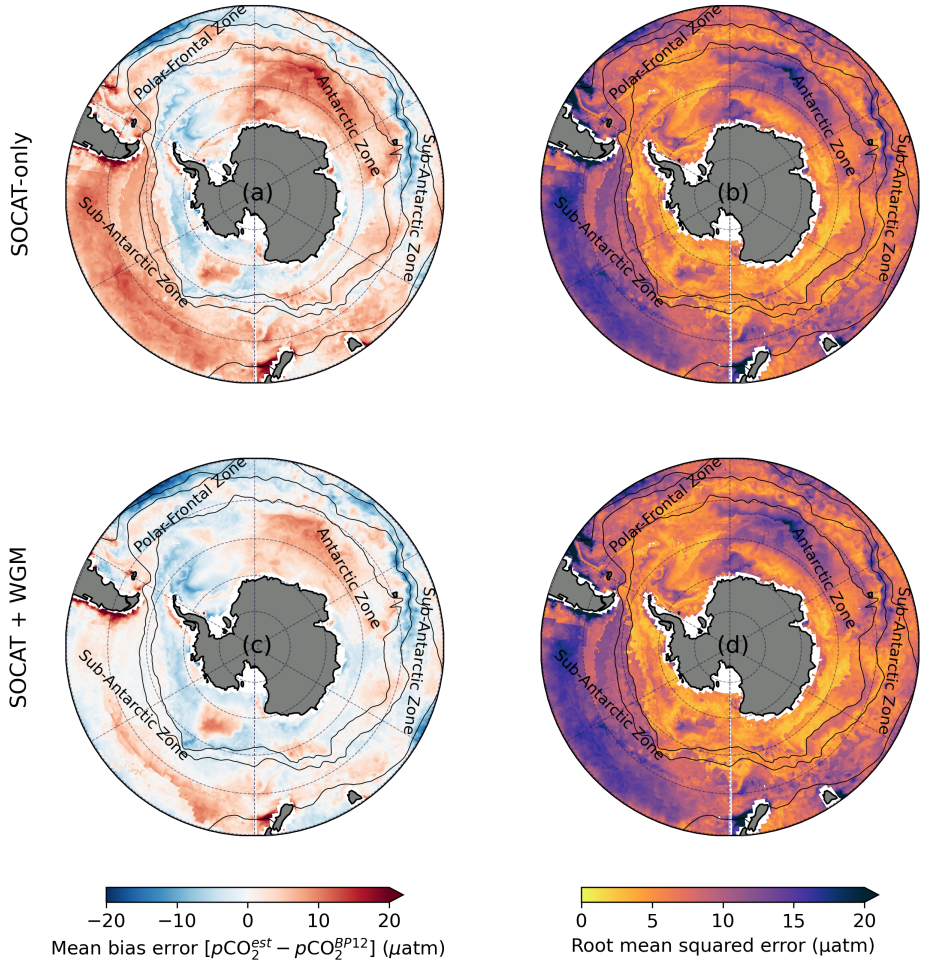


Figure 4.13: The Southern Ocean (south 40°S) maps of the mean bias errors (first column) and root mean squared errors (second column) of ensemble average pCO_2 estimates. The top row corresponds to the SOCAT-only scenario while the bottom row corresponds to the SOCAT + WGM scenario.

In summary, higher Pearson correlations with lower RSME scores (cf. Table 4.4) indicate that the SOCAT + WGM scenario improves the pCO_2 reconstruction uncertainties in the entire Southern Ocean and its regions compared to the SOCAT-only scenario as expected. However, how much uncertainty improvement got achieved from SOCAT-only to SOCAT + WGM? To answer this question, a quantification of pCO_2 mapping improvements due to the addition of winter data in the Southern Ocean was achieved with the estimated percentage change of the RMSE scores. For an initial RMSE_i score (from SOCAT-only) and a final RMSE_f score (from SOCAT + WGM), this measure is calculated as follows:

$$\%\text{Change} = \frac{\text{RMSE}_f - \text{RMSE}_i}{\text{RMSE}_i} \times 100 \tag{4.3.1}$$

A distinguishing characteristic of %Change is that its calculation does not account for context; that is, the assumptions underlying the experimental setting. A positive value will imply a percentage increase while a negative will mean a percentage decrease. The results of this analysis are summarized in Table 4.5.

Table 4.5: A summary of the percentage change of the RMSE scores from SOCAT-only to SOCAT + WGM

Mapping approach	%Change (in %)			
	SO	SAZ	PFZ	AZ
FBM-FNN	-39.85	-48.86	-16.52	-32.81
SCR-FNN	-46.27	-40.26	-54.92	-50.97
MKM-FNN	-50.24	-51.6	-48.93	-53.02
ENS	**-36.59**	**-46.69**	**-35.04**	**-27.59**

Table 4.5 shows that all the percentage change calculations resulted in negative values, meaning that they all characterize a percentage decrease in RMSE scores from the SOCAT-only scenario to SOCAT + WGM. These results show that regardless of the mapping approaches used, the RMSE scores get improved significantly in the whole Southern Ocean and regionally when adding winter data to the summer-biased underway SOCAT ship observations. Regionally, for example, the largest improvement is recorded in the SAZ when using the ENS mapping approach (46.69%).

4.3.5 The reconstructed pCO_2 seasonal cycle

Besides analysing the mapping errors of the two OSSE sampling scenarios SOCAT-only and SOCAT + WGM to evaluate the impacts of clustering-based methods and investigate the sampling biases, the mean amplitude seasonal cycle estimates were also analysed (hereinafter referred to as "anomalies"). As the dominant mode of variability in the Southern Ocean, the seasonal cycle has often been used on many occasions as a diagnostic framework to assess performance (Mongwe et al., 2018; Gregor et al., 2018; Gloege et al., 2021). Here, diagnosing the anomalies of the mean seasonal cycle estimates (Figure 4.14) allows us to further evaluate the impact of the seasonal sampling biases, particularly in winter, on the reconstructed pCO_2 across methods and Southern Ocean regions.

In the SOCAT-only scenario, all the methods highlighted that there is a significant winter bias on pCO_2 mean seasonal cycle estimates (Figure 4.14a-d). However, the influence of this winter bias slightly differs between mapping methods and regions. For example, except in the polar frontal zone (PFZ) where the FMB-FNN approach resolved almost all the seasonal biases (Figure 4.14c), larger seasonal cycle (SC) anomalies of pCO_2 were observed in the Southern Ocean, the sub-Antarctic zone (SAZ), the Antarctic zone (AZ) as shown in Figure 4.14a, b, and d, respectively. SC anomalies larger than 20 μatm were observed in the SAZ for the FMB-FNN and MKM-FNN approaches (Figure 4.14 This is likely due to lower observation-density areas such as the Pacific Ocean sector (Figure 4.1).

Considering the second OSSE sampling scenario, SOCAT + WGM, the SC anomalies of reconstructed pCO_2 show almost no signs of seasonal sampling biases regardless of the gap-filling approach used (Figure

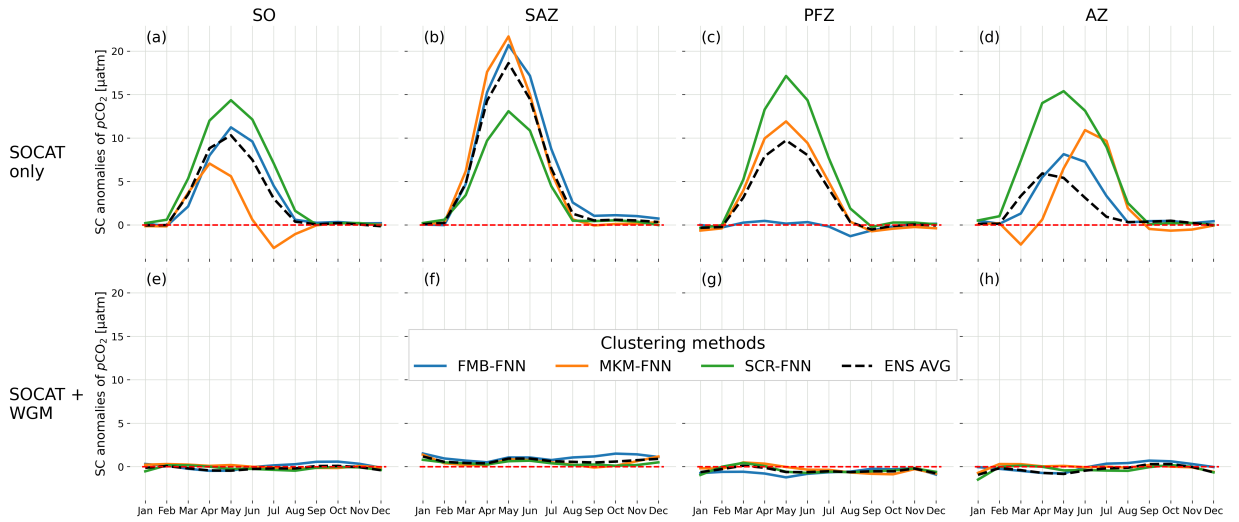


Figure 4.14: Anomalies of the mean seasonal cycle (SC) of the reconstructed pCO_2 using FMB-FNN, MKM-FNN, SCR-FNN, and ENS AVG. The top row, panels **(a)**, **(b)**, **(c)** and **(d)**, shows the zonal mean SC anomalies corresponding to the SOCAT-only scenario experiments while the bottom row, panels **(e)**, **(f)**, **(g)** and **(h)**, shows the zonal mean SC anomalies corresponding to the SOCAT + OWC scenario experiments. The red dashed line corresponds to the zero line. Abbreviations in the figure legend correspond to SO = Southern Ocean; SAZ = sub-Antarctic zone; PFZ = polar frontal zone; and AZ = Antarctic zone.

4.14e-h). Overall, in this scenario, all the mapping approaches performed equally well in reproducing the seasonal variability of the surface ocean pCO_2 across the Southern Ocean and its regions.

4.3.6 The reconstructed meridional gradients of pCO_2

The influence of the seasonal sampling biases in the Southern Ocean extends to the meridional gradients of surface ocean pCO_2 in the region, whose estimates can be used as an evaluation metric of the two OSSE scenarios and the seasonal sampling biases. High-latitude regions (e.g.: south of 60°S, see Figure 4.15) such as the north of the AZ and south of the PFZ are known to be areas of intense winter out-gassing of carbon modulated by the upwelling of Circumpolar Deep Waters (CDWs). However, it can be seen in Figure 4.15 that in both SOCAT-only and SOCAT + WGM scenarios, this process got weakened in summer and reversed (around 60°S) during the seasonal poleward evolution of summer productivity. This is likely associated with the retreat of the marginal ice zone because the meridional variation of the surface ocean pCO_2 results from its effect on the sea surface temperature (SST), and hence the CO_2 solubility (Wu et al., 2019).

The upwelling phenomenon, whose global significance has previously been overlooked (Wu et al., 2019), is very crucial in shaping the spatial distribution of the surface ocean pCO_2 in the Southern Ocean. Given the seasonal and meridional changes of this process in the region (DeVries et al., 2017), it was important to look at the influence of winter sampling gaps on meridional gradients of surface ocean pCO_2. Compared to SOCAT-only, the SOCAT + WGM scenario significantly improves estimates of the meridional gradient of pCO_2, particularly in mid-latitude areas as shown in Figure 4.15a-b. The canonical seasonal cycle of pCO_2

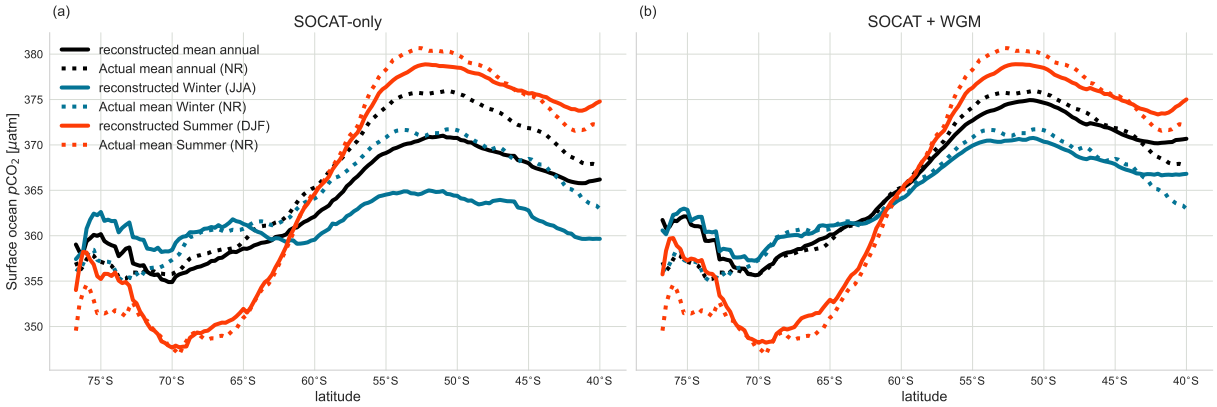


Figure 4.15: Meridional gradient of the surface ocean pCO_2 from ensemble average (ENS) estimates, where the solid curves depict the reconstructed means in Winter (in blue), Summer (in red) and mean annual (in black), while the dashed curves depict their respective actual means calculated from the Nature Run pCO_2. Panels **(a)** and **(b)** correspond to the two experimental sampling scenarios: SOCAT-only and SOCAT + WGM. Abbreviations in the figure legend correspond to NR = Nature Run; DJA = December January February; and JJA = June July August.

in the Southern Ocean is mainly governed by the seasonal cycle of SST (Takahashi et al., 2009; Landschützer et al., 2015; Gruber et al., 2019). This governance is something observations and data products and also SST-driven ocean models such as BIOPERIANT12 (Section 2.1) strongly agree on in the Southern Ocean (Figure 3.2), particularly in mid- and high-latitude areas also called upwelling areas (DeVries et al., 2017; Mongwe et al., 2018; Hauck et al., 2020).

4.4 Discussions

Alone, the Southern Ocean is thought to account for a significant portion of the contemporary ocean uptake of anthropogenic CO_2 emissions (Landschützer et al., 2015; Friedlingstein et al., 2021), but air-sea CO_2 flux estimates in this region are uncertain mainly due to surface ocean pCO_2 estimates which are based on scarce and sparse observations that are strongly biased toward summer (Gray et al., 2018; Bushinsky et al., 2019). The discussion below interprets the results presented above based on the two observational system scenarios: SOCAT-only and SOCAT + WGM, detailed in Section 4.2.2.

4.4.1 Simulated SOCAT in-situ observations in OSSE

The density of observations from underway SOCAT ships is not temporally and spatially homogenous in the Southern Ocean (Figure 4.1). However, in practice, such sampling limitations and biases in the region influence important variability features such as the seasonal cycle which can be resolved synthetically through data simulation when running OSSEs. On the other hand, when using surface ocean CO_2 measurements from the SOCAT database, it is important to be aware that these in-situ observations from underway ships are not collected at the ocean surface (Gregor et al., 2019). For example, the in-situ temperatures that coincide with CO_2 data in the SOCAT database are thus different from the sea surface temperature (SST) products

that are used to estimate the surface ocean pCO_2 and then calculate air-sea CO_2 fluxes (Bakker et al., 2016; Goddijn-Murphy et al., 2015). The inconsistency in in-situ and remotely sensed temperature results in a theoretical difference between pCO_2 measured at the ship intake depth and the surface due to warming and cooling (Takahashi et al., 2009). For years now, this ocean-surface biasedness in in-situ observations has been a topic of major consideration in the marine CO_2 observations community.

Due to the use of synthetic observations in OSSEs, such a temperature correction does not need to be applied, which is an advantage this study has. However, with real-world observations, the potential impacts of this discrepancy cannot be neglected (Gregor et al., 2019). Therefore, one important caveat to be considered in this study is that the ocean-surface biasedness of SOCAT in-situ observations is corrected during the data simulation and assimilation stages on the Nature Run. On the other hand, in the Southern Ocean, the Coupled Model Intercomparison Project version 5 (CMIP5) Earth system models used in climate predictions disagree on the phasing of the seasonal cycle of the surface ocean pCO_2 (Lenton et al., 2013; Mongwe et al., 2018). Nicholson et al. (2016) and Monteiro et al. (2015) argued that this poor representation of pCO_2 stemmed from the failure to resolve sub-mesoscale or fine-scale processes due to the coarse resolutions of the models.

The Nature Run in the OSSE system of this study is the outputs of a mesoscale-resolving ocean model (cf. Section 4.2.3), which likely is one of the possible best representations of the processes of interest in this study although it remains uncertain whether they are properly captured by the BIOPERIANT12 ocean model configuration (Section 2.1). Either way, this justifies the use of this ocean model output in the OSSEs to evaluate clustering-based methods and investigate the implications of winter sampling gaps (existing in in-situ observations from underway SOCAT ships) on pCO_2 reconstruction in the whole Southern Ocean. As a model resolving the sub-grid scale dynamics in the Southern Ocean, the BIOPERIANT12 ocean model is assumed to represent the scales of variability of interest. However, the implications of the imperfection of BIOPERIANT12 in terms of not getting the phasing of the seasonal cycle of pCO_2 suggested by existing reconstructions (Figure 3.2), remains unknown.

4.4.2 Necessity of clustering-regression in the Southern Ocean

Due to the dynamic of the surface ocean pCO_2 in the Southern Ocean, which is regionally complex with a seasonal cycle largely influenced by the variability of the sea surface temperature and other processes (Takahashi et al., 2009), clustering choice of the Southern Ocean is essential in order to improve modelling of CO_2 in the region. Particularly, combining this clustering with regression methods such as the feed-forward neural network (FNN) provides us with a more dynamic and regional understanding of ocean productivity, which is more based on underlying physical and biogeochemical drivers rather than only climatological values (Thomalla et al., 2011; Fay and McKinley, 2014; Gregor and Gruber, 2021). The characterization of

Fay and McKinley (2014) biomes (Figure 4.8a) considered these physical and biogeochemical drivers for the surface ocean pCO_2 as well as the creation of mini-batch K-Means clusters (Figure 4.8c). The seasonal cycle by being the dominant mode of variability in the Southern Ocean (Mongwe et al., 2018), knowledge of the seasonal variability is also necessarily relevant to clustering of the domain before applying any regression methods.

The seasonal cycle reproducibility (SCR) of pCO_2 allowed the creation of three main seasonal regions based on the intensity of the SCR: lower, medium and higher (Figure 4.8b). The response of the surface ocean pCO_2 to the underlying physics and biogeochemistry of different seasonal regions resulted in an additional classification of the zonal regions based on the existence of intra-seasonal variability, and seasonal or annual signals. This zonal and seasonal regionalisation (Figure 4.8b) contributes toward a better understanding of the differences that the clustering step makes. It improves the estimates of the Southern Ocean's sensitivity to climate forcing and potentially allows a more robust analysis of the effects of long-term climate trends (Thomalla et al., 2011).

Furthermore, using the mini-batch K-Means clustering method with various feature combinations and the number of clusters as shown in Figure 4.9 resulted from testing a range of 3 to 25 clusters. This process leveraged two clustering metrics (Silhouette score and Inertia or sum of squared distances/errors) (Rousseeuw, 1987) to measure the performance of each clustering configuration (Figure 3.8), instead of using the test score of the regressions in the next step like in Gregor et al. (2019). An advantage of doing this was to help maintain a better balance between the size of a cluster and the data density (cf Figure 4.8c). Using the test score of the regression step as a more complete indicator of performance (Gregor et al., 2019) might have introduced a representational bias if clusters were imbalanced.

Moreover, it is practically very difficult to train regression machine-learning models on high-resolution ocean model data in a large domain such as the whole Southern Ocean without clustering first the domain. This was less difficult in Chapter 3 because a small sub-domain of the Southern Ocean was used instead (Section 4.2.3), hence the clustering step was not necessary. Therefore, this study on such high-resolution ocean model data in the whole Southern Ocean is a novelty. From a viewpoint of required computing power, for example, training the FNN regression algorithm on these high-resolution data points in separate clusters is not only far much faster than running it in the entire Southern Ocean but also demands CPU and GPU processing powers that the Centre of High-Performance Computing (CHPC) was able to provide us from their computing facilities. With such facilities, running an OSSE test for each one of the three clustering-regression approaches (Section 4.2.3) took about 96 hours (4 days), which resulted in about 12 days for all 17 FNN regression models from all the clusters.

4.4.3 Differences in methods and their impacts

A succinct performance analysis of different clustering methods, particularly their out-of-sample scores (cf. Section 4.3.3), shows that they provide uniquely different outcomes (Table 4.3). While evaluating them through the SOCAT-only and SOCAT + WGM scenarios, findings reveal how much spatial and temporal limitations in in-situ pCO_2 observations influence the mapping outcomes of the surface ocean pCO2 in the Southern Ocean. Therefore, differences in results of FMB, MKM, and SCR methods stem from how each one handles data-limitation challenges in space and time in the whole Southern Ocean and regionally. Out of the 17 clusters generated for this study (Section 4.3.2), 4 clusters corresponded to FMB, 6 to SCR, and 7 to MKM. While the same FNN regression algorithm was applied separately across all 17 clusters, the training and validation resulted in 17 different FNN models. They all differed from each other with their underlying optimal architectures, particularly with the number of hidden layers, and the number of neurons/nodes in each hidden layer.

In general, empirical methods involved in the estimation of the surface ocean pCO_2 are not yet very suitable for making predictions in time (Rödenbeck et al., 2015; Keppler and Landschützer, 2019). According to the authors, these methods are just gap-filling approaches, therefore, are best suitable for interpolations in the study domain. However, one of the advantages of using this OSSE framework (Section 4.2.3) is that all data from the observing system simulation are allocated to the training and validation of the FNN regression algorithm (in the tow-step gap-filling methods) while the rest of the data from Nature Run, unseen during the training and validation, are used to thoroughly test the reconstruction skills of the resulting optimal models and compare them (Table 4.3).

4.4.4 Quest for a robust pCO_2 estimation

By the term "robust", I mean something that is in good condition, close to perfect with lower bias. Thus, having robust estimates in empirical modelling increases the confidence level in various findings. A few recent studies applied novel statistical and machine learning (ML) techniques to obtain robust estimates of ocean carbon sink from pCO_2 estimates using in-situ surface ocean CO_2 observations (Landschützer et al., 2014; Gregor et al., 2019; Denvil-Sommer et al., 2019; Gregor and Gruber, 2021). From the necessity of clustering-regression approaches to the differences in these approaches with their impacts on the modelling outcomes in the Southern Ocean, the discussion is now on the adoption of the ensemble-average approach as a robust solution to the quest in the presence of multiple expert methods.

From applying the ensemble-average approach which incorporated the MKM, SCR and FMB learning outcomes into an ensemble average for the final estimation of surface ocean pCO_2, it was found that the ensemble-average estimates were more robust than results from individual methods in the whole Southern

Ocean and across regions (cf. Tables 4.3 and 4.4). In other words, the overall performance of the ensemble average is improved, which provides robust estimates of surface ocean pCO_2, although computing an ensemble might likely weaken the geographical meaning of the ensemble domains or clusters. The robustness of the ensemble-average estimates here implies significant improvements in the mapping performances, which results in better reconstructions of surface ocean pCO_2 than using the individual methods. However, this might also be due to the reduced aliasing in the clustering process that constitutes the first step in each of these mapping approaches, especially with the SCR learning approach. Introducing a clustering technique based on the seasonal cycle reproducibility (SCR) of surface ocean pCO_2 in the Southern Ocean is a novelty on its own compared to Gregor et al. (2019) and Denvil-Sommer et al. (2021) who mainly focused on Fay and McKinley (2014) biomes which do not account the seasonal cycle. On the other hand, the existing relationships between the surface ocean pCO_2 and proxy variables likely improve in clusters because of the data-size increase, especially during winter in the SOCAT + WGM scenario.

Furthermore, Gregor et al. (2019) reported the convergence of pCO_2 reconstructions scores to a common threshold and spoke of "hitting the wall" in terms of improving pCO_2 reconstruction uncertainties and biases. In their study, the authors made use of mini-batch K-Means and (Fay and McKinley, 2014) approach for clustering their study domain. Based on our findings, however, including in the ensemble clustering method such as the SCR of surface ocean pCO_2 would likely get us beyond the "wall" in the real world where learning models is achieved using a finite set of training datasets, which makes the learning an ill-posed problem. The purpose of ensemble learning is thus to construct learners and combine them in an intelligent way to obtain an overall robust model. In this study, results suggest that ensemble learning results are robust and more accurate than results from any of the single clustering methods in the ensemble. Therefore, this increases confidence in the study findings in the Southern Ocean.

4.4.5 Influence of winter sampling biases

Underway SOCAT ship measurements of surface ocean CO_2 are essential to global and regional mappings of the surface ocean pCO_2 using machine learning algorithms (Hauck et al., 2020). However, these underway SOCAT ship measurements in the Southern Ocean are very scarce and far much denser in summer than in winter (cf. Figure 4.1), therefore, introducing seasonal biases in observations (Gray et al., 2018). In this section, the discussion mainly makes use of the ensemble results.

Using the seasonal cycle anomaly (Figure 4.14) as a mode to diagnose the impacts of seasonal sampling biases in the reconstructed pCO_2, contrasting regional biases and uncertainties were found in surface ocean pCO_2 estimates in the whole Southern Ocean and regionally. In the SOCAT-only scenario, these regional contrasts are particularly evident in the sub-Antarctic zone (SAZ) (cf. Figure 4.14a-c) which is largely influenced by winter biases in the Pacific Ocean sector (cf. Figure 4.13a). In addition, this may also be

modulated by the biological carbon pump in the SAZ (Thomalla et al., 2011) and the seasonal cycle of the sea surface temperature (Takahashi et al., 2009) which are seasonally intensified during summer (Mongwe et al., 2018).

Overall, the SAZ is by far the largest contributor to mapping errors for all methods and across regions (Tables 4.3 and 4.4), with ENS achieving an RMSE score of 10.87 µatm in the SOCAT-only scenario and 5.80 µatm in the SOCAT + WGM scenario (Table 4.4), therefore, capturing the high variability of the seasonal cycle in the region (Figure 4.14d-f) as also reported by Bushinsky et al. (2019). The observance of these larger errors in the SAZ relative to PFZ and AZ may result from a combination of two factors: (1) the variability of surface ocean pCO_2 is largest in this zonal region with short spatial and temporal decorrelation length-scales, especially in the Pacific Ocean sector (Gregor et al., 2018); (2) the available proxy variables may not adequately represent the physical and biogeochemical processes in the region (Gregor et al., 2017). On the other hand, the anomalies of the mean seasonal cycle of pCO_2 in the Southern Ocean as a whole may be dominated by those of the PFZ in the SOCAT-only scenario and the AZ in the SOCAT + WGM scenario (Figure 4.14).

Furthermore, assuming the feasibility of winter gap mitigation along underway SOCAT ship tracks from summer scheduled voyages led us to the SOCAT + WGM scenario which attempted to resolve the winter sampling biases. Results suggested that this would contribute to reducing uncertainties and biases in contemporary pCO_2 reconstructions of at least 36.59% for the whole Southern Ocean. With the largest regional improvement in the SAZ (46.69%), this finding is also confirmed by the regional anomalies of the mean seasonal cycle (cf. Figure 4.14d-f) for which the mitigation of the winter sampling bias would reduce the mean SC anomalies to ± 2.5 µatm. Observation-based reconstructions of the mean state of surface ocean pCO_2 are thus very sensitive to resolving the seasonal variability. Based on this finding, a poorly constrained seasonal variability in in-situ observations may be contributing to existing differences and uncertainties in data products in the Southern Ocean (Hauck et al., 2020; Friedlingstein et al., 2021).

Therefore, the lack of observations covering the full seasonal cycle in the Southern Ocean challenges gap-filling methods, which will result in noisy reconstructions of surface ocean pCO_2 and disagreements between different methods. This finding is supported by many studies such as Rödenbeck et al. (2015) and Gregor et al. (2017). The importance of improving winter observations is in part related to the quality of prior information about the seasonal cycle of surface ocean pCO_2, which is likely worse in the Southern Ocean compared to most other ocean regions. Bushinsky et al. (2019) also came to a similar conclusion while conducting a reassessment of air-sea CO_2 flux estimates in the Southern Ocean with additional profiling float observations. The authors also tested the influence of including measurements at the times and locations of SOCCOM float observations.

4.4.6 SOCAT-only vs SOCAT + WGM

Recalling that to address this problem the study considered two simulation scenarios SOCAT-only and SOCAT + WGM, now the question is how do these two scenarios compare to each other in terms of improving the out-of-sample RMSE scores of the clustering methods? In other words, how do seasonal sampling gaps in the Southern Ocean (Bakker et al., 2016; Bushinsky et al., 2019) affect observation-based reconstructions of surface ocean pCO_2 in the region? In Figure 4.16, I make a graphical representation of RMSE scores in Table 4.3 to facilitate the discussion.

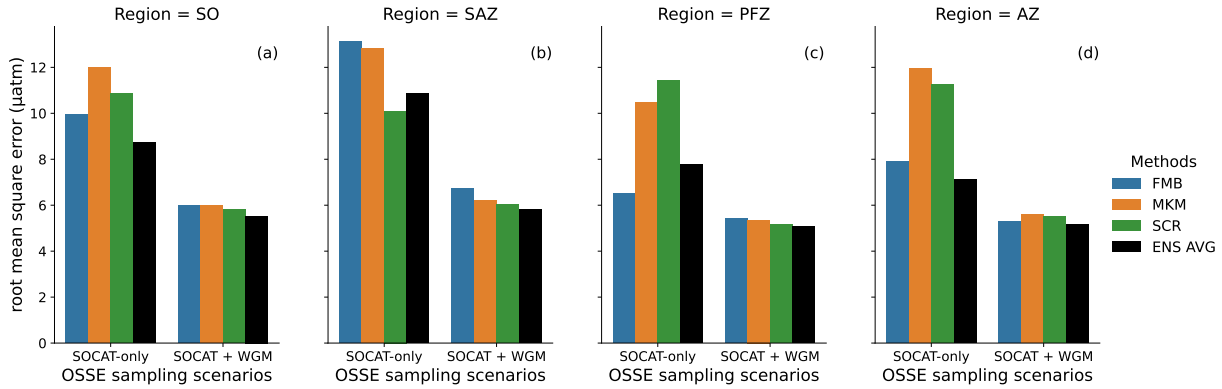


Figure 4.16: Zonal assessment of the cluster-based reconstruction of pCO_2 using RMSE as the main performance metric. The abbreviations are SO = Southern Ocean; SAZ = sub-Antarctic zone; PFZ = polar frontal zone; and AZ = Antarctic zone.

The SOCAT-only scenario simulated the real-world underway SOCAT ship observations which are biased toward summer using 2011-2019 ship tracks in the Southern Ocean. And to compensate for winter sampling gaps present in this OSSE scenario, a simulated sampling was run in winter along these SOCAT-ship routes in the Southern Ocean. Data from this simulation called winter gap mitigation (WGM) was then added to the initial SOCAT-only data to compensate for the seasonal sampling biases due to these observational gaps in winter, hence, the second OSSE scenario SOCAT + WGM. Using the regional comparison framework (Section 2.5), the reconstruction uncertainty from both OSSE scenarios, graphically summarized in Figure 4.16 indicates the importance of resolving winter observational gaps in the whole Southern Ocean and regionally, regardless of the methods used.

Despite some challenges due to winter data gaps in the SOCAT-only scenario, all the mapping methods outperformed by far in the SOCAT + WGM scenario where winter data gaps were filled with winter observations along a few realistic SOCAT-ship routes. Compared to SOCAT-only, the SOCAT + WGM scenario found significantly lower RMSE scores in the whole Southern Ocean and regionally (Figure 4.16). In other words, the increasing performance in the SOCAT + WGM scenario resulted in significant improvements in the surface ocean pCO_2 reproduction throughout the Southern Ocean as shown in Figure 4.13 (RMSEs and MBEs) and confirmed by the estimates of the percentage change in RMSE scores (Table 4.5). Therefore,

RMSE scores reducing drastically after mitigating the winter observational gaps (Figure 4.15) is likely a function of the importance of improving winter observations of surface ocean CO_2 in the Southern Ocean. This highlights the importance of sampling scales that pCO_2 observing systems used in the whole Southern Ocean and regionally.

Furthermore, it was found in Chapter 3 that the seasonal scale variability and meridional gradient matter the most and hypothesized that the underway SOCAT ship data in the Southern Ocean resolve the meridional gradients for summer only due to winter sampling gaps in shipboard observations in the region. Here, this hypothesis is validated using surface ocean pCO_2 estimates from the two OSSE scenarios (Figure 4.15). Moreover, the canonical seasonal cycle in the Southern Ocean presents a strong in-gassing and out-gassing of surface ocean CO_2 in summer and winter, respectively (Takahashi et al., 2009; Gregor et al., 2019). Notwithstanding phasing differences, Figure 4.15a highlights that during the in-gassing in summer, pCO_2 levels reach a minimum (< 350 µatm) likely due to biological production which draws down the inorganic carbon to the surface. As for the out-gassing in winter, pCO_2 levels reach a maximum (> 362.5 µatm) resulting from the deep-mixing delivery of carbon-rich waters to the surface caused by the upwelling of CDWs in high-latitude regions (Takahashi et al., 2009; DeVries et al., 2017).

However, what would happen if the synthetic underway ship routes in winter were chosen differently? Given the complexity of the Southern Ocean due to its remoteness and extreme weather conditions, underway SOCAT ship observations in the region result only from scheduled voyages which predominantly take place in summer. These scheduled voyages for the Southern Ocean hydrographic surveys coordinated through research programs such as the Global Ocean Ship-based Hydrographic Investigation Program (GO-SHIP) are carried out on limited and prescribed routes. Therefore, the SOCAT + WGM scenario by itself is an extreme sampling case in the Southern Ocean. This makes an alternative choice of underway ship routes for synthetic winter sampling in the region very unrealistic. However, with recent technological advances, for example, autonomous unmanned surface vehicles (USVs) or robots such as Saildrones could be designed and deployed in a way to accompany ships through the sampling of ocean areas where underway ships are unable to reach.

4.5 Conclusions

In this chapter, an OSSE framework was developed and used to evaluate the clustering methods in two-step pCO_2 reconstructions in the Southern Ocean and their contribution to the bias and root mean square error as a measure of uncertainties. This OSSE system used a high-resolution (daily $\times \pm$10km) NEMO-PISCES forced ocean model output for my synthetic truth or Nature Run. Then, two sampling experiments or OSSE scenarios including SOCAT-only and SOCAT + WGM were considered. The SOCAT-only scenario

simulated from Nature Run the underway SOCAT ship observations which are biased toward summer. In the other scenario, a simulation called winter gap mitigation (WGM) was added to SOCAT-only to compensate for the seasonal sampling biases due to winter observational gaps, hence the acronym SOCAT + WGM. Afterwards, the impacts of seasonal sampling biases stemming from underway SOCAT ship observations in the region were investigated. The OSSE system used three two-step gap-filling approaches engineered around a state-of-the-art machine-learning algorithm: the feed-forward neural network (FNN). The first approach was FMB-FNN as its first step was based on Fay and McKinley (2014) biomes (FMB). The second was MKM-FNN referring to its first step built after the unsupervised ML algorithm called mini-batch K-Means (MKM). Lastly, the third method was SCR-FNN whose first step was based on the seasonal cycle reproducibility (SCR) of pCO_2 in the Southern Ocean. The two research questions in Section 4.1.2, asked of these experiments are answered as follows.

1. **Which one of the clustering methods used offers greater gains through the two-step reconstruction approach?**

The clustering choice is very important in two-step pCO_2 reconstructions in the whole Southern Ocean and regionally. Comparing the three clustering methods, it was found that the results differed regionally depending on two sampling simulation scenarios including SOCAT-only and SOCAT + WGM. For example, considering the SOCAT-only scenario in the Southern Ocean as a whole, FMB outperformed MKM which was outperformed by SCR, while in the two SOCAT + WGM scenario, SCR outperformed both FMB and MKM throughout the Southern Ocean.

Findings show that while the clustering learning methods of two-step approaches perform adequately throughout the Southern Ocean, the ensemble average learning (ENS) combining the three approaches together improves the results significantly and makes findings more robust and less clustering-method specific. Therefore, this suggests that using ENS is the optimal balance between cluster-based approaches in terms of robustly reducing biases and uncertainties in pCO_2 reconstructions in the whole Southern Ocean and regionally.

The use of SCR clustering in this study is not only a novelty but also is the only method out of the three clustering methods that truly considers the seasonal cycle (SC) of surface ocean pCO_2 before clustering the Nature run field, knowing that SC is the dominant mode of variability in the Southern Ocean.

2. **How is the contribution of these clustering methods to uncertainties of the reconstruction sensitive to the contrasting summer and winter sampling scenarios?**

The magnitude of improvement is enormous if ships could sample during winter in the Southern Ocean along the underway SOCAT ship tracks operational in summer. Based on objective estimates using ensemble average learning, the percentage decrease in RMSE scores (characterizing uncertainties) from the

SOCAT-only scenario to SOCAT + WGM is generally around 30% for most of the regions, up to 47% for the sub-Antarctic zone (SAZ). The anomalies of mean seasonal cycle estimates in the Southern Ocean revealed that if data are added in winter (e.g.: SOCAT + WGM), then the biases would disappear. With the paucity of sampling in the Pacific Ocean sector of the SAZ, resolving winter data gaps as characterized by SOCAT + WGM would not only lead to the SAZ as the region with the largest improvement ($\sim 47\%$) but also reduce regional anomalies of the mean seasonal cycle estimates to about ± 2.5 µatm. This shows how reconstructions of surface ocean pCO_2 are very sensitive to resolving seasonal variability and therefore, highlights the importance of sampling scales that observing systems adopt in the whole Southern Ocean and regionally.

Multiple runs of idealized experiments through OSSEs support that a poorly constrained seasonal variability in in-situ observations contributes to discrepancies and uncertainties in data products in the Southern Ocean (Williams et al., 2017; Hauck et al., 2020; Friedlingstein et al., 2021). The results of this chapter based on the use of a simulated observational system show that resolving winter sampling biases in shipboard CO_2 measurements in the Southern Ocean can potentially improve pCO_2 reconstructions in the region, up to the point of reducing to about zero the seasonal cycle anomalies. This is however reliant on a substantial increase in the number of winter voyages in the Southern Ocean. This emphasises the importance of appealing to autonomous observing systems to improve coverage of surface ocean CO_2 observations in the Southern Ocean, as it will be investigated in the next chapter.

Chapter 5

Getting over the wall: a scale-sensitive integrated pCO_2 observing system for the Southern Ocean

5.1 Introduction

The existence of seasonal and intra-seasonal sampling biases in current and new observing systems is a significant cause of the convergence of the pCO_2 reconstruction performances to a relatively common value – proverbially known as "hitting the wall" (Gregor et al., 2019). This problem is mainly due to the limits of operational pCO_2 reconstructions when actual observations are either sparse or biased in time and space, in the Southern Ocean (Bushinsky et al., 2019; Bakker et al., 2016). In Chapter 3, I showed that observing both the seasonal cycle of meridional gradients and the intra-seasonal variability of surface ocean pCO_2 $\left(pCO_2^{ocn}\right)$ is crucial to significantly reduce biases and uncertainties in pCO_2 reconstructions. While underway SOCAT ships mostly sample the Southern Ocean meridionally, they do so with a strong summer seasonal bias, which impacts the uncertainties and biases of pCO_2 reconstructions in the region (Gray et al., 2018; Gregor et al., 2018; Djeutchouang et al., 2022). In a sub-domain of the model in Chapter 3 and the whole Southern Ocean in Chapter 4, these winter observational gaps were mitigated by repeating underway SOCAT ship measurements on the same cruise tracks used in summer using synthetic model data. This resulted in enormous improvements of the mean seasonal cycles throughout the whole Southern Ocean and regionally by reducing significantly biases and uncertainties in surface ocean pCO_2 reconstructions (see Figures 4.13 and 4.14; Tables 4.4 and 4.5). However, synthetic observations were used at a semi-idealized winter effort equivalent to the summer historical sampling effort. One major problem is that, in the real world, this is practically difficult in winter with underway ships. Therefore, in this chapter, the most realistic approach is explored based on autonomous Unmanned/Uncrewed Surface Vehicles (USVs).

Notwithstanding the contribution of autonomous sampling platforms to resolving the seasonal cycle and intra-seasonal variability in the Southern Ocean, I showed in Chapter 3 that simultaneously observing the meridional gradients of the surface ocean pCO_2 at high temporal resolution is not operationally possible using Wave Gliders and floats. Djeutchouang et al. (2022) showed that the main limitation with floats is the 10-day sampling period as well as a quasi-Lagrangian zonal sampling (Figure 3.5; Table 3.2). On the other hand, Wave Gliders, which can directly sample the pCO_2 at the surface ocean at high frequency and with great accuracy (Sutton et al., 2021; Monteiro et al., 2015), operate in the Southern Ocean as pseudo

stationary moorings (Monteiro et al., 2015; Nicholson et al., 2022).

Using multiple runs of semi-idealized Observing System Simulation Experiments (OSSEs), the USV Saildrones (Sutton et al., 2021) stood out to be the optimal candidate capable of simultaneously achieving high resolution of both meridional gradients and the seasonal cycle compared to Wave Gliders and Profiling floats. However, the OSSEs were only conducted in a strategically selected high-EKE sub-domain at the east of the Atlantic Ocean sector of the Southern Ocean (Section 3.2.2). Given that this experimental sub-domain was representative of the scales of variability of interest, the finding with USV Saildrones was hypothesised to apply to the Southern Ocean as a whole.

5.1.1 Aims and Questions

This chapter aims to propose a scale-sensitive sampling strategy that addresses the limitations of underway SOCAT ship observations in the whole Southern Ocean and strengthens the existing summer-biased ship-based reconstructions of surface ocean pCO_2 in order to "get over the wall" in terms of reconstruction uncertainties and biases. The validity of the USV-Saildrone conjecture in Chapter 3 is thus investigated through the integration of the USV-Saildrone observations. This tests the impacts of zonal and meridional sampling strategies, adjunct to SOCAT-like ship sampling, on the pCO_2 reconstruction in the Southern Ocean. To achieve these aims, the OSSE framework developed in Section 4.2.3 is implemented and the ensemble average learning is used for surface ocean pCO_2 estimates. Thus, the key questions this chapter is addressing are the following:

1. Do simultaneous high-resolution observations of the seasonal cycle of the meridional gradients of pCO_2^{ocn}, which USV Saildrones are capable of providing, significantly improve the biases and uncertainties in pCO_2 reconstructions in the Southern Ocean as a whole? In other words, do the key findings from Chapter 3 extend to the Southern Ocean as a whole?

2. Would meridional sampling with USV Saildrones in the Southern Ocean (as a whole) significantly improve the currently used zonal circumpolar strategy?

5.2 Methodology

Here I offer a brief overview of pseudo-observations or synthetic data from the two simulated observing systems resulting from the Observing System Simulation Experiments (OSSEs) run in this study. First, underway SOCAT ships already perform regularly gridded measurements to estimate surface ocean pCO_2 (Bakker et al., 2016). Second, there are USV Saildrones which have the possibility not only to be equipped with new sensors to provide physical and biogeochemical measurements (Sutton et al., 2021) but also to simultaneously perform high-resolution observations of meridional gradients and the seasonal cycle. In

this section, we also recall the regression approach for reconstructing the surface ocean pCO_2 according to various simulated sampling scenarios, and how it is used in the OSSEs. This includes the semi-idealized modelling experiments used to evaluate the impacts of integrating SOCAT shipboard measurements with the new observing system (USV Saildrone sensors) on operational pCO_2 reconstructions when actual observational data are scarce and sparse in space and time in the Southern Ocean (cf. Section 4.2.2 and Figure 4.1).

5.2.1 OSSEs in the Southern Ocean

Simulations of SOCAT-shipboard sampling follow the underway sampling system on board volunteering observing ships contributing to the SOCAT dataset. The resulting simulated data is thus sparsely limited in time and space like in the SOCAT dataset itself in the Southern Ocean, which is biased toward summer (Gray et al., 2018). This corresponds to pseudo-observations derived from the SOCAT-only scenario used in Chapter 4. Specifically, the simulation process behind these pseudo-observations of SOCAT ships is described in Section 4.2.2.

Synthetic USV-Saildrone circumpolar sampling

The autonomous Unmanned Surface Vehicle (USV) Saildrone has many key features such as its higher speed and use of wind- and solar-powered energy to extend its endurance (Meinig et al., 2019; Whitt et al., 2020), especially in ocean regions with harsh low light winter conditions like the Southern Ocean (Figure 5.1). Practically, a USV Saildrone can observe the mole fraction of CO_2 (xCO_2) in both the surface ocean and atmospheric boundary layer which can be paired with ancillary variables such as the sea surface salinity (SSS), sea surface temperature (SST) (Sutton et al., 2021). Figure 5.1 illustrates the first USV-Saildrone circumnavigation of Antarctica in just about 7 months providing high-resolution zonal (hourly) measurements. Exploiting these advantages of USV Saildrones, synthetic USV-Saildrone observations in the Southern Ocean are simulated by sub-sampling the surface ocean pCO_2 and ancillary variables from the BIOPERIANT12 ocean model output (Section 2.1).

Zonal and meridional pseudo-observational scenarios

In simulating the USV Saildrones, two sampling options are considered, both in which I assume that the USV Saildrones sample the full seasonal cycle including summer and winter. The first option, denoted by zUSV, is to sample the Southern Ocean zonally along the Antarctic Circumpolar Current (ACC), which follows the actual zonal strategy (Figure 5.1b) adopted by the USV Saildrone in Sutton et al. (2021). This option is subdivided into two scenarios: zUSV[1] and zUSV[3], which correspond to the simulated zonal USV sampling with respectively one Saildrone deployment (Figure 5.2a) and three simultaneous USV Saildrone deployments (Figure 5.2b) to circumnavigate Antarctica like the 2019 Antarctica circumnavigation with

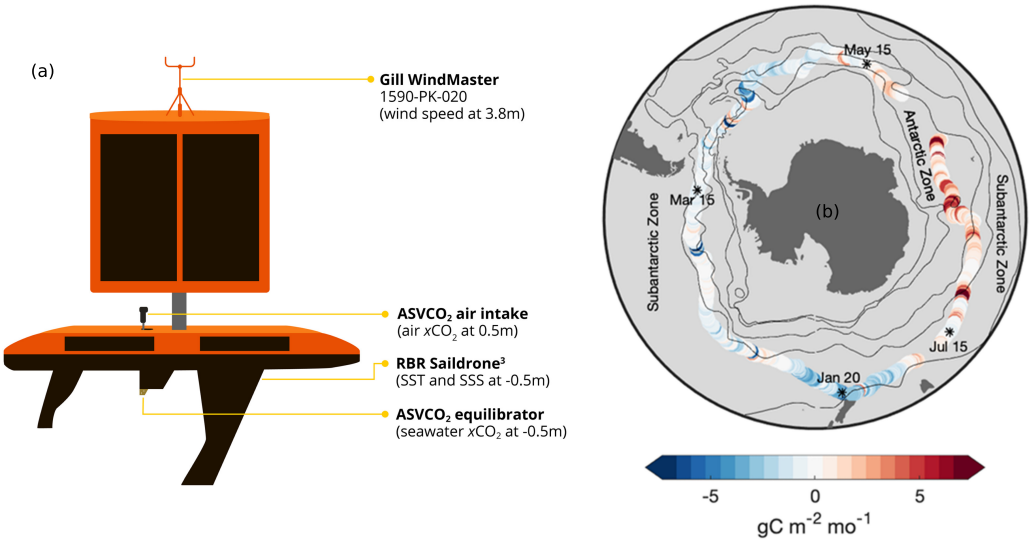


Figure 5.1: The Unmanned Surface Vehicle (USV) Saildrone (adapted from Sutton et al. (2021)). **(a)** schematic diagram of the USV Saildrone (with the location of a few sensors) deployed in the Southern Ocean in 2019. **(b)** sampling tracks of the USV Saildrone from Sutton et al. (2021) on which are shown the air-sea CO_2 fluxes calculated based on USV-Saildrone-measured ΔxCO_2. Dates and ⁕ illustrate the locations of the USV Saildrone with time as it completed in 198 days (about 7 months) the first circumnavigation of the Southern Ocean. Black lines indicate climatological locations of the major Southern Ocean fronts from Orsi et al. (1995).

the USV Saildrone described by Meinig et al. (2019). The zUSV[3] scenario performs a total of 194 545 pseudo-observations from the three simultaneous zonal USV deployments (Figure 5.1b) while zUSV[1] makes 63 145.

The second option, denoted by mUSV, adopts a meridional sampling strategy to sample the Southern Ocean between 40°S and 60°S with the main assumption that the USV Saildrone is operationally capable in the region. Thus, four potential deployment ports are considered as depicted in Figure 5.1: (1) Cape Town in the Atlantic Ocean (CPT), (2) Reunion in the Indian Ocean (RIN), (3) Hobart in the Pacific Ocean (HBT), and (4) Auckland in the Pacific Ocean (APC). This second option corresponds to simultaneous sampling with four USV Saildrones deployed separately from the four potential deployment ports (Figure 5.2c), which makes in total 114 558 pseudo-observations. Thus, this second option assumes five meridional sampling lines: mUSV[CPT], mUSV[RIN], mUSV[HBT], mUSV[APC], and mUSV[4].

Considering the two sampling scenarios zUSV[3] and mUSV[4] (Figure 5.2b-c) allows also to increase spatial data coverage in the Southern Ocean as recommended by Lenton et al. (2013) and Monteiro et al. (2010, 2015). These USV-Saildrone sampling experiments enable us to test the sensitivity of the pCO_2 reconstructions to these orthogonal sampling strategies (Figure 5.2c).

In each of the sampling scenarios described in this section, a year-round output of the high-resolution (daily $\times \pm$10km) NEMO-PISCES forced ocean model (BIOPERIANT12) is sub-sampled at various sampling tracks (Figure 5.2) of USV Saildrones to create pseudo-observations of surface ocean pCO_2 and its proxy variables. Given the daily temporal resolution of the BIOPERIANT12 output, these pseudo-

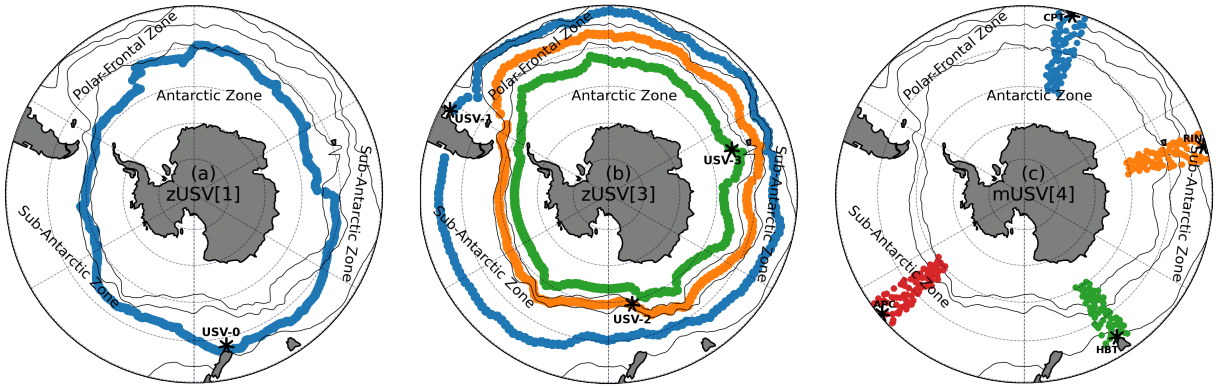


Figure 5.2: Simulated sampling tracks of the USV-Saildrone deployments in the Southern Ocean. Panels **(a)** and **(b)** correspond to the two scenarios zUSV[1] and zUSV[3] of the zonal USV sampling strategy. Panel **(c)** shows the five meridional USV sampling scenarios mUSV[CPT], mUSV[RIN], mUSV[HBT] and mUSV[APC], plus mUSV[4]. Black stars (⁕) correspond to potential ports of the USV deployment while the black lines indicate climatological locations of the major fronts from BIOPERIANT12 model data calculated as in Orsi et al. (1995).

observations are also daily.

5.2.2 Integrated sampling systems

Seven combined sampling systems based on Figure 5.2 are considered and tested in this study: SOCAT + zUSV[1], SOCAT + zUSV[3], SOCAT + mUSV[CPT], SOCAT + mUSV[RIN], SOCAT + mUSV[HBT], SOCAT + mUSV[APC], and SOCAT + mUSV[4]. In these multi-platform observational systems, by SOCAT we mean the simulated SOCAT shipboard measurements (biased toward summer) or simply referring to the SOCAT-only scenario in Chapter 4. These integrated sampling systems have a few coverages in winter due to the year-round sampling by simulated USV Saildrones.

This multi-platform integration is realistic because it uses a closer representation of the sampling extension of the 2019 Antarctica circumnavigation with the USV Saildrone(Meinig et al., 2019; Sutton et al., 2021). Specifically, SOCAT shipboard measurements essential to the machine-learning reconstruction of air-sea CO_2 fluxes (Landschützer et al., 2016; Gregor et al., 2019; Hauck et al., 2020), ship-based measurements of surface ocean CO_2 are thus likely to continue as an opportunistic sampling on scheduled re-supply voyages in Antarctica (Gregor et al., 2017). Additionally, observing platforms such as floats and moorings typically rely on ships for their deployment (Talley et al., 2019).

5.2.3 Ensemble learning regression

The ensemble average learning (ENS) is the three-member ensemble of clustering-regression methods where the clustering step consists of Fay and McKinley (2014) biomes (FMB), the seasonal cycle reproducibility (SCR) clusters, and the mini-batch K-Means (MBKM) clusters. The regression step consists of the feed-forward neural network (FNN) algorithm. More details of this FNN regression are given in Section 2.3.2

while details of the clustering methods are given in Section 4.2.4.

Based on Chapter 4's results (e.g.: cf. Figure 4.16), the ENS approach outperformed all the individual methods and derived estimates were found to be robust compared to estimates from single methods. Therefore, in this chapter, ensemble-averaged learning is used to perform the regression of surface ocean pCO_2 against ancillary variables.

5.2.4 OSSE framework synthesis

This section recalls key stages of the OSSE framework/system used in this study. This OSSE system is similar to the framework used in Chapter 4 (see Section 4.2.3) differing with the addition of a new observing system, the USV Saildrone, to supplement the SOCAT-only scenario. This OSSE framework works as follows:

1. A forced and mesoscale resolving (daily $\times \pm$10km) ocean model run of the Southern Ocean (BIOPERIANT12) is used as the "ground truth" for the system, termed as Nature Run (see Section 2.6).
2. The Nature Run field output is used to extract "pseudo-observations" co-located in time and space with the SOCAT and the newer Saildrone (USV) pCO_2 observing systems.
3. The synthetic observations are used in the three-member ensemble-learning approach (ENS) made up of the two-step machine-learning methods used in Chapter 4 (see Section 4.2.3).
4. The pCO_2 reconstructions are made with the resulting model and compared with the Nature Run to quantify the improvements from the additional OSSEs (e.g.: zUSV and mUSV).

In the following sections, the root mean squared error (RMSE) described in Section 2.3.6 is used as the primary assessment metric. It also serves as a measure of uncertainty in the pCO_2 reconstructions. In some circumstances, Pearson's correlation coefficient (r) and bias are also used in combination with the RMSE score. The reported scores will be out-of-sample scores, calculated on data that were not involved in simulations of pseudo-observations used in the model training, making these mapping scores adequate to assess the various OSSE-integrated sampling systems.

5.3 Results

It is important to recall that results from the SOCAT-only run from OSSEs in the previous chapter (see Sections 4.3.3 and 4.3.4) serve as the main reference in the analysis of the results of this chapter. This OSSE run is equivalent to the real-world use of the seasonally biased (toward summer) SOCAT data in the Southern Ocean (Bakker et al., 2016) for reconstructions of the surface ocean pCO_2 in the region (Gregor et al., 2018; Bushinsky et al., 2019; Gloege et al., 2021; Denvil-Sommer et al., 2021). The results presented

here are from the OSSE framework (see Section 5.2.4) whose sampling scenarios are described in Section 5.2.1. The reconstruction of surface ocean pCO_2 results from using the ensemble average approach (ENS) in Section 5.2.3. The reconstructed surface ocean pCO_2 from multi-platform systems (Section 5.2.2) shall also be referred to as a product, for example, the product SOCAT + zUSV[1] (SOCAT supplemented by the single zonal USV) is simplified to zUSV[1] product. In this formalism z and m refer to zonal and meridional respectively.

5.3.1 Impacts on the seasonal cycle of pCO_2

Figure 5.3 shows the anomalies of the mean seasonal cycle (SC) of surface ocean pCO_2 estimates resulting from the SOCAT + zUSV[1], SOCAT + zUSV[3] and SOCAT + mUSV[4] integrated OSSEs. The reconstructed SC biases due to the winter sampling gaps from SOCAT data in the Southern Ocean are still visible in the first multi-platform integrated system (Section 4.3.5).

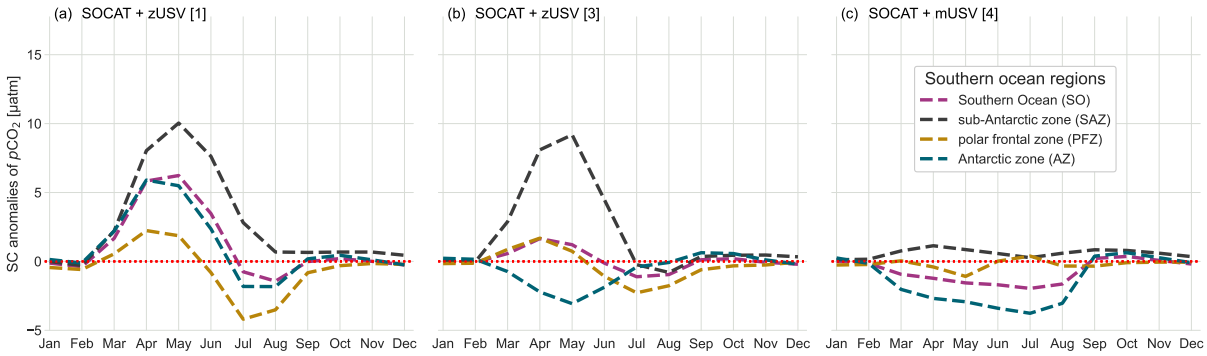


Figure 5.3: Anomalies of the seasonal cycle (SC) of the reconstructed pCO_2 from the ensemble average (ENS) approach. Panels **(a)**, **(b)** and **(c)** show the regional mean SC anomalies corresponding to multi-platform integrated sampling systems SOCAT + zUSV[1], SOCAT + zUSV[3] and SOCAT + mUSV[4]. The red dashed line corresponds to the zero line. They show that the reconstructed bias weakens significantly as the meridional gradient is increasingly better resolved.

Throughout the Southern Ocean, the SOCAT + zUSV[1] system results in higher SC anomalies as shown in Figure 5.3a, particularly during the winter months. In this OSSE scenario, the SC anomalies in the whole Southern Ocean seem to be predominantly modulated by the AZ anomalies. However, they all have a similar phasing throughout the Southern Ocean with the highest and smallest magnitudes in the SAZ and PFZ, respectively. Following the same zonal sampling strategy, Figure 5.3b shows that when increasing the meridional data coverage by adding three zonal distinct deployments, the reconstructed SC anomalies drop down throughout the Southern Ocean. This results in a significant reduction of biases in the SC in winter months, particularly in the PFZ and AZ regions, and the Southern Ocean as a whole.

In the SAZ, however, the SOCAT + zUSV[3] pCO_2 product in winter months yielded large SC anomalies that were approximatively equivalent to the SC anomalies yielded by the SOCAT + zUSV[1] pCO_2 product (Figure 5.3b).

5.3.2 Sensitivity of the meridional gradients of pCO_2 to OSSE scenarios

Figure 5.4 shows the reconstructed means of the meridional gradients of surface ocean pCO_2 from combined pCO_2 observing systems SOCAT + zUSV[1] and SOCAT + zUSV[3], and SOCAT + mUSV[4]. The meridional variability of summer, winter and annual means show significantly different behaviours, particularly for high latitudes around 65-75°S.

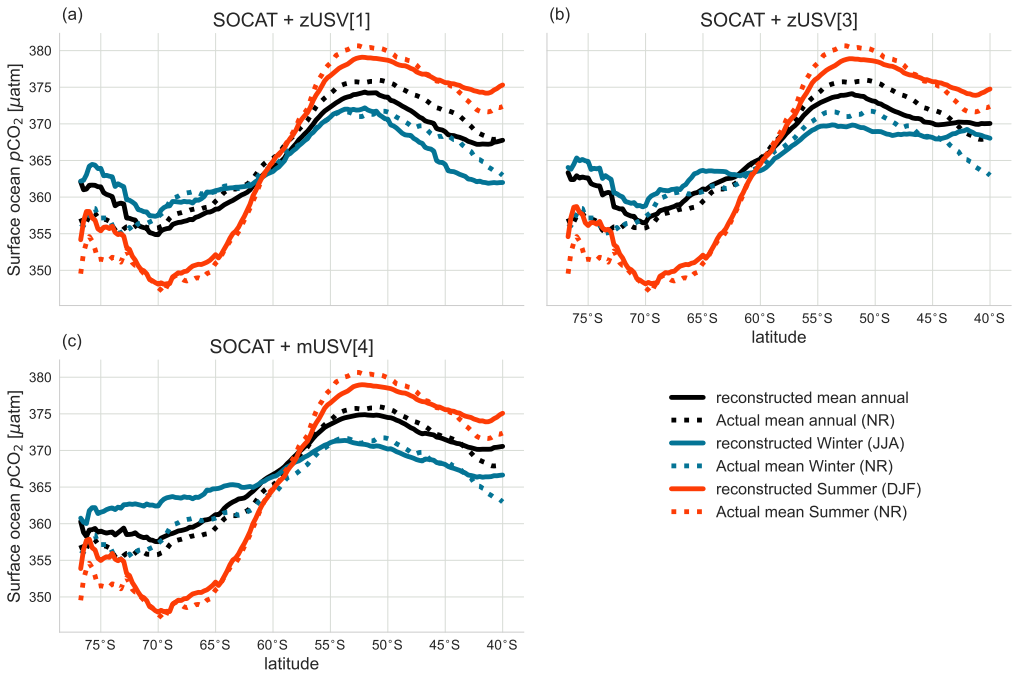


Figure 5.4: The meridional gradients of the surface ocean pCO_2 from the multi-platform integrated sampling systems. Solid curves depict the reconstructed means in Winter (in blue), Summer (in red) and mean annual (in black), while the dashed curve depicts in each case the actual mean calculated from the Nature Run pCO_2. Abbreviations in the figure legend correspond to NR = Nature Run; DJA = December January February; and JJA = June July August.

The interpretation of these meridional gradients focuses on changes in their phasing and magnitude. Thus, it can be seen in Figure 5.4 that the meridional gradients in summer (December-February) are equally reconstructed across all the OSSE scenarios as expected. This implies that the meridional gradient in summer is not sensitive to adding USV-Saildrone observations to underway SOCAT ship data. Therefore, the SOCAT-only scenario being biased toward summer, the combination with USV-Saildrone observations does not add much information in summer as the underway ship-based observations already resolve the meridional gradients as shown in Chapter 3.

However, adding the USV-Saildrone observations has significantly different implications on the meridional gradient in winter (June-August) across all the OSSE scenarios around 40-55°S and 65-75°S (Figure 5.4). The results are equally sensitive for the SOCAT + zUSV[1] and SOCAT + zUSV[3] products except in the mid-latitudes (about 40-55°S) where the former reveals a decrease of about 3.5 µatm. In the SOCAT

+ mUSV[4] scenario, a higher impact on the reconstructed meridional gradient in winter is observed in the mid-latitudes and high-latitudes with the former highlighting the intense winter out-gassing of carbon from the upwelling of carbon-rich deep waters towards the surface with subsequent CO_2 release.

Furthermore, compared to the meridional gradient from other OSSE scenario products, this addition also improves the reconstructed mean annual of the meridional gradient with the largest improvement resulting from the SOCAT + mUSV[4] experiment (Figure 5.4c).

5.3.3 Spatial distribution of pCO_2 biases

The maps of mean bias error (MBE) between actual pCO_2^{ocn} from the Nature Run and the reconstructed pCO_2^{ocn} reveal a significant difference in the distribution between each multi-platform integrated system (Figure 5.2). Somehow, this is to be expected given that these systems use different sampling strategies with different data coverages.

It can be seen that the pCO_2^{ocn} MBEs improve regionally with the increase in the sampling effort of multi-platform systems (Figure 5.5). The pCO_2^{ocn} MBEs in the SOCAT + zUSV[1] product are largely higher in the north of the Pacific basin of the SAZ ($>$ 10.5 µatm) and in a few areas south of the Atlantic and Indian basins of the AZ compared to other regions. In these areas, particularly the north of the Pacific basin, SOCAT-based in-situ observations of surface ocean CO_2 are very sparse in space and time as shown in Figure 4.1. Figure 5.5b shows that increasing the spatial coverage of data with a similar zonal sampling using three USV Saildrones (SOCAT + zUSV[3]) results in a slight reduction of MBEs in these data-sparse areas. However, the performance difference between the two systems SOCAT + zUSV[1] and SOCAT + zUSV[3] seems very small throughout the Southern Ocean except in the SAZ and AZ.

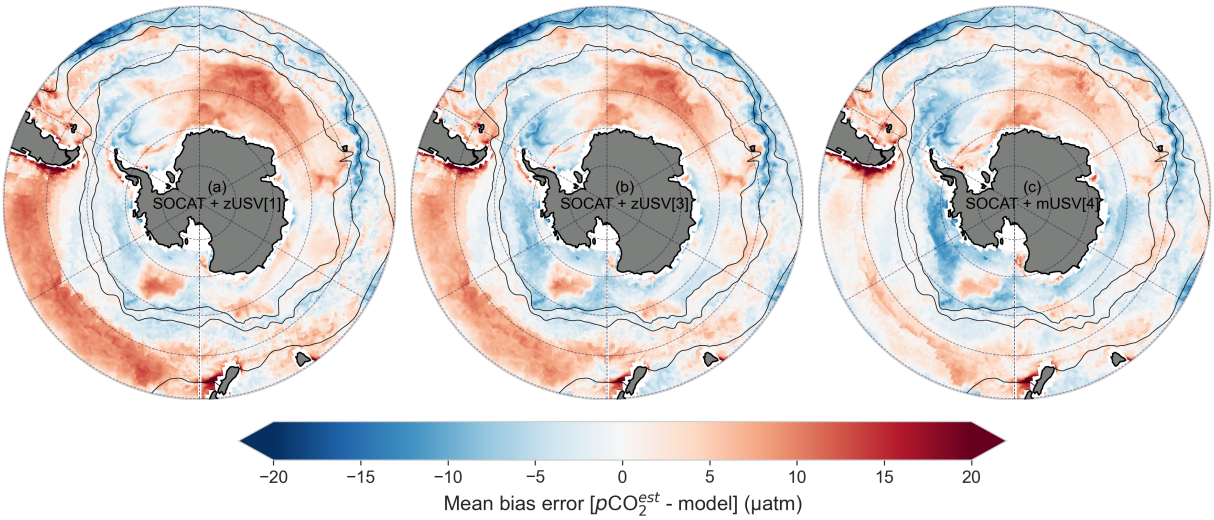


Figure 5.5: Maps of mean bias errors (MBEs) of pCO_2^{ocn} reconstructions from the three major integrated sampling systems SOCAT + zUSV[1], SOCAT + zUSV[3] and SOCAT + mUSV[4] shown in panels **(a)**, **(b)** and **(c)**, respectively.

Figure 5.5 also shows that there are areas which can be identified as error hot spots, such as coastal areas

of the south of Chile and New Zealand, the Pacific sector of the SAZ and the Indian sector of the AZ that show positive pCO_2^{ocn} MBEs (Figure 5.5). While SOCAT + mUSV[4] weakens spatial positive MBE anomalies, the negative anomalies seem to strengthen in the East Pacific and West Atlantic (Figure 5.5c). However, despite pCO_2^{ocn} MBEs being steadily larger in these coastal areas for all the multi-platform systems (Figure 5.5), the SOCAT + mUSV[4] system significantly reduced these pCO_2^{ocn} MBEs throughout the Southern Ocean (Figure 5.4c). The improvements with the increase in the sampling effort are remarkably significant in these error hot spots, particularly in the Pacific and Indian basins, as compared to the circumpolar sampling approach. This is likely due to the meridional sampling strategy (Figure 5.2c) adopted by USV Saildrones in this OSSE scenario, enabling USVs to sample meridionally like underway SOCAT ships in regions such as the Pacific where ship operations are almost non-existent in winter.

The positive biases in the Atlantic and Indian oceans south of the Polar front suggest that the area has some seasonal dynamics that may be particularly sensitive to sampling biases because these positive biases weaken significantly from zUSV to mUSV. This points to a sampling sensitivity which is mostly addressed by the mUSV OSSE. These areas of high positive biases are in the sub-tropics and the seasonal cycle of the Net Primary Production in these areas is reversed.

Analysis of the actual surface ocean pCO_2 data from the Nature Run output and the estimated surface ocean pCO_2 using the ensemble-averaged approach (Section 5.2.3) indicates that the Pacific Ocean sector of the SAZ is the basin with the highest mean bias errors. The annual mean of the surface ocean pCO_2 in the SAZ for this product was computed at 369.41 ± 11.35 µatm, whereas the actual mean from Nature run output was 374.65 µatm. The combined SOCAT + zUSV[1] and SOCAT + zUSV[3] products converge a bit more with estimates of 371.83 ± 10.24 µatm and 372.55 ± 9.46 µatm, respectively, while the SOCAT + mUSV[4] product converges even much better with the annual mean pCO_2 estimate of 374.05 ± 8.49 µatm.

5.3.4 Ensemble-learning performance

Regression scores

Table 5.1 shows the performance metrics of the ENS approach. Using a zonal sampling strategy, results show that the OSSE-integrated sampling systems SOCAT + zUSV[1] and SOCAT + zUSV[3] allowed to achieve respectively the RMSE scores of 7.10 µatm and 7.04 µatm in the Southern Ocean. This suggests that SOCAT + zUSV[3] outperformed SOCAT + zUSV[1] with a small RMSE margin of ~ 0.06 µatm. Regionally, this RMSE margin is higher in the SAZ (~ 2.2 µatm) and PFZ (~ 0.22 µatm) whereas in the AZ the SOCAT + zUSV[1] outperform SOCAT + zUSV[1] with the RMSE margin of 0.22 µatm. Overall, zonally sampling the Southern Ocean by circumnavigating Antarctica using an integrated sampling system of more than one USV Saildrone contributes slightly better to reducing uncertainties in pCO_2 reconstructions than using one USV. However, in regions such as the AZ, it would be sufficient to only use one USV.

Table 5.1: A summary of the performance estimates of the ENS method run with various OSSE-integrated sampling systems. The root mean squared error of surface ocean pCO_2 is defined for the Southern Ocean (SO) and then all the zonal regions as defined by Lovenduski et al. (2008) while the bias or mean bias error and Pearson's correlation are computed for the whole SO. The best results for each column are shown in bold according to each sampling strategy. Values in the table are significantly different from the mean for the column and clustering method, with a p-value (p) < 0.05 for the two-tailed Z test.

Sampling strategy	OSSE-integrated sampling scenarios	Root mean squared error (µatm)				Bias (µatm)	r
		SO	SAZ	PFZ	AZ		
Zonal	SOCAT + zUSV[1]	7.10	8.51	5.78	**6.06**	**0.41**	0.78
	SOCAT + zUSV[3]	**7.04**	**6.37**	**5.56**	7.60	0.93	0.80
Meridional	SOCAT + mUSV[CPT]	7.59	7.47	6.17	7.97	0.83	0.75
	SOCAT + mUSV[RIN]	7.53	8.98	7.30	**6.30**	1.83	0.77
	SOCAT + mUSV[HBT]	11.55	7.66	5.74	14.9	3.99	0.69
	SOCAT + mUSV[APC]	10.72	6.15	5.66	14.07	3.48	0.71
	SOCAT + mUSV[4]	**6.34**	**5.97**	**5.15**	6.70	**0.48**	**0.82**

Considering the meridional sampling strategy against the zonal sampling, individual systems achieved relatively large biases, RMSE and r scores with respective average differences of about 2.5 µatm, 9.3 µatm and 0.72 µatm in the Southern Ocean (Table 5.1). The combined OSSE-integrated sampling system SOCAT + mUSV[4] scores consistently low for all three metrics (Table 5.1). However, there are two remarkable results from the SOCAT + mUSV[HBT] and SOCAT + mUSV[APC] that show the highest uncertainties in the Southern Ocean pCO_2 reconstruction with RMSEs of 11.55 µatm and 10.72 µatm, respectively. These highest uncertainties seem to have been contributed by the uncertainties from the AZ region in which SOCAT + mUSV[HBT] and SOCAT + mUSV[APC] achieved RMSE scores of larger magnitudes of 14.4 µatm (Table 5.1) compared to 6.5 µatm in the SAZ and PFZ. This reflects the nature of the Pacific basin where the AZ is smaller compared to the Atlantic and Indian basins (cf. Figure 5.2). Therefore, accounting also for the sea ice extension, the USV Saildrones can only sample small areas of the AZ along meridians as illustrated by the APC and HBT tracks in Figure 5.2c.

Henceforth, further analyses focus more on the two OSSE-integrated sampling scenarios with the larger data coverage in the Southern Ocean. This includes the circumpolar and meridional cases SOCAT + zUSV[3] and SOCAT + mUSV[4] for which zUSV[3] and mUSV[4] added to SOCAT data 194 545 and 114 558 pseudo-observations, respectively. Comparatively, the meridional sampling strategy with SOCAT + mUSV[4] achieved in the Southern Ocean a lower RMSE score (6.34 µatm) than the zonal sampling strategy with SOCAT + zUSV[3] achieving an RMSE score of 7.04 µatm. Although their biases are slightly different in the Southern Ocean, the result with RMSE scores is also regionally consistent and confirmed by the Pearson correlation in the Southern Ocean as a whole. Based on this result, the meridional integrated sampling system SOCAT + mUSV[4] outperformed the zonal integrated sampling system SOCAT + zUSV[3] throughout

the Southern Ocean.

Quantifying the changes in RMSE scores

Besides analysing the performance of the multi-platform sampling systems, I am also interested in how much change these systems bring to pCO_2 mapping uncertainties in the Southern Ocean. Ensemble-averaged results from the sampling systems were evaluated based on the RMSE score between reconstructed and true pCO_2 as a way to determine whether adding USV-Saildrone observations to SOCAT data decreased differences from true Nature Run output. Thus, the percentage change (Equation 4.3.1) in pCO_2 RMSE scores was calculated to better understand the impact that combining sampling systems has on reconstructed surface ocean pCO_2. The results of this analysis are summarized in Table 5.2; negative values correspond to a percentage decrease in the RMSE score while positive values correspond to a percentage increase.

Table 5.2: Summary of the percentage change (in %) of uncertainty scores due to multi-platform sampling integration throughout the Southern Ocean.

Sampling strategy	OSSE-integrated sampling scenarios	Root mean squared error (μatm)			
		SO	SAZ	PFZ	AZ
Zonal	SOCAT + zUSV[1]	-19.5	-22.21	-25.99	-16.41
	SOCAT + zUSV[3]	-20.18	-41.77	-28.81	4.83
Meridional	SOCAT + mUSV[CPT]	-13.95	-31.72	-21.0	9.93
	SOCAT + mUSV[RIN]	-14.63	-17.92	-6.53	-13.1
	SOCAT + mUSV[HBT]	30.95	-29.98	-26.5	105.52
	SOCAT + mUSV[APC]	21.54	-43.78	-27.53	94.07
	SOCAT + mUSV[4]	-28.12	-45.43	-34.06	-7.59

Adopting the current zonal sampling strategy (Figure 5.2a-b), decreasing RMSE scores were recorded as shown in Table 5.2 where adding zUSV[1] and zUSV[3] outputs have a broadly similar impact in the Southern Ocean as a whole except in the AZ region where the error slightly increases for the zUSV[3] output (4.83%). When a meridional sampling strategy was used in the Southern Ocean as a whole, the addition of USV-Saildrone observations (i.e. mUSV[4]) reduced the regional RMSE scores in the SOCAT + mUSV[4] product by 0.55 to 4.97 µatm (Tables 4.4 and 5.2). Relative to the mean RMSE score across all regions of 6.04 µatm, this reduction is somewhat very large and resulted in an average improvement of the pCO_2 RMSE score of 28.8% (Table 5.2).

5.4 Discussions

5.4.1 Multi-platform integration

The success of the multi-platform integration seats on three levels: observing platforms (ships and USV Saildrones, Section 5.2.1), clustering methods and regression models as respectively first and second steps in two-step pCO_2 reconstructions (Section 4.2.4). A brief overview discussion on why and how I got here is offered as follows.

The key aim of this chapter is to propose and discuss a scale-sensitive sampling strategy that addresses the limitations and strengthens the existing summer-biased ship-based reconstructions of pCO_2 in order to "get over the wall". In a comparative global reconstruction, Gregor et al. (2019) suggested that we, the community, may have reached the limits of the ability of gap-filling methods to reduce biases and uncertainties in pCO_2 estimates. The authors hypothesized that the reason why uncertainties kept prevailing was most likely due to basin-scale observational gaps as well as sampling aliases in mesoscale-intensive ocean regions such as the Southern Ocean. This hypothesis could not be rejected in Chapter 3 where I showed the importance of scale-sensitive sampling in reducing uncertainties using meridional, intra-seasonal, and single-ship line winter samplings approaches in mesoscale- and eddy kinetic energy (EKE)-sensitive sub-domain of the Southern Ocean (Djeutchouang et al., 2022).

To extend the testing of this scale-sensitive sampling hypothesis to the Southern Ocean as a whole, I make use of a two-step (clustering-regression) reconstruction approach. The clustering step is necessary in this approach because when applied at a very large scale, machine learning algorithms struggle to represent accurately pCO_2 unless spatial coodinates are included as a proxy (Gregor et al., 2017, 2019). I thus found in Chapter 4 that the ensemble of clustering methods including a novel supervised clustering approach based on the seasonal cycle reproducibility of pCO_2 was very sensitively effective in reducing bias and root mean square error. Therefore, how does this help to "get over the wall"?

Southern Ocean pCO_2 data from underway ship observations in SOCAT are very biased toward summer (Figure 4.1b). The chosen multi-platform sampling systems (Section 5.2.2) have seasonal coverages in austral winter months and have achieved lower RMSE scores (Table 5.1) compared to the SOCAT-only scenario in Chapter 4 (cf. Table 4.4). These findings suggest that the multi-platform integration improved the reconstruction skill of the surface ocean pCO_2 throughout the whole Southern Ocean and regionally. Thus, I am also interested here in how the winter biases are reduced by the proposed sampling networks. To do this, both zonal and meridional sampling strategies of USV Saildrones are analyzed and discussed.

5.4.2 Effect of zonal vs meridional sampling by USVs on winter biases

The current sampling strategy used in the Southern Ocean by autonomous observing platforms is zonal because of their Lagrangian nature (Majkut et al., 2014; Mazloff et al., 2018; Bushinsky et al., 2019) and so is the sampling strategy of the recently developed observing platform USV Saildrone (Meinig et al., 2019; Sutton et al., 2021). However, one of the features of this USV Saildrone is its capability for a circumpolar sampling of the Southern Ocean (see Section 5.2.1), hence the two zonal sampling scenarios zUSV[1] and zUSV[3] (Figure 5.2a-b). On the other hand, in the mUSV[4] simulations (Figure 5.2c), I leveraged one of the USV Saildrone strengths (see Section 5.2.1) to extend the meridional sampling strategy (proposed in Chapter 3) to the Southern Ocean as a whole.

The large-scale patterns with positive mean bias errors highlighted in the Pacific and Indian sectors of the SAZ and AZ, respectively (Figure 5.5), were initially revealed by the spatial distribution of the surface ocean pCO_2 biases and uncertainties from the SOCAT-only product (Figure 4.13a-b). Results reveal that these large-scale patterns of deviation are improved as USV-Saildrone observations are added to the SOCAT-only data. However, the improvement rate was lower when using a zonal sampling strategy than using a meridional sampling. For example, when considering the RMSE scores (Tables 4.4 and 5.1) in the Southern Ocean as a whole, the margin estimates are 1.6 µatm, 1.68 µatm and 2.38 µatm for the SOCAT + zUSV[1], SOCAT + zUSV[3] and SOCAT + mUSV[4] products, respectively, the higher the margin the better the OSSE. Additionally, Figure 5.5 shows that large-scale patterns of deviation get smaller in the SOCAT + mUSV[4] product, which is also confirmed in Figure C.2. Therefore, these findings are consistent with the hypothesis on the importance of resolving the seasonal cycle of the meridional gradient and because of that the SOCAT + mUSV[4] system leads to a large reduction in biases and uncertainties in pCO_2 reconstructions in the Southern Ocean.

Sampling the Southern Ocean using autonomous observing platforms such as USV Saildrones has the advantage to cover a full seasonal cycle which is lacking from underway SOCAT ships whose surface ocean pCO_2 observations have very few data points in the winter months (Figure 4.1b). When the data coverage is increased with up to 3 circumpolar USV Saildrones (Figure 5.1b), I found that the resulting integrated system (SOCAT + zUSV[3]) achieved almost a similar impact as using only one USV Saildrone deployed zonally (SOCAT + zUSV[1]), where the percentage decreases in RMSE scores were estimated at 20.18% and 19.5%, respectively (cf. Tables 5.1 and 5.2). This suggests that, in the case of the zonal sampling strategy, fewer USV Saildrones are required to achieve a similar result if machine learning methods are used. However, error hot-spots such as the Pacific sector of the SAZ and the Indian sector of the AZ, are large sources of uncertainties in mean seasonal cycles of pCO_2, especially in winter months (cf. Figure 5.3a-b).

Similarly, by increasing the data coverage with a meridional sampling strategy, I found that this allows us to achieve far better results throughout the Southern Ocean than using a single USV Saildrone deployed meridionally (Tables 5.1 and 5.2; Figures 5.3 and 5.5). For example, with up to 4 meridional deployments (SOCAT + mUSV[4]) but with 79 987 data points as less as using 3 zonal deployments (SOCAT + zUSV[3]), the SOCAT + mUSV[4] system outperformed SOCAT + zUSV[3] in reducing winter anomalies in mean seasonal cycle reconstructions (Figure 5.3b-c) and also improved better the mean annual and winter of the reconstructed meridional gradients of surface ocean pCO_2 .

The analysis in Section 5.3.3 demonstrates that biases in the SAZ, for example, were predominantly due to existing observational gaps in the Pacific basin, which are worse in winter months (see Figure 4.1). This indicates that pseudo-observations are well consistent within the ability of the gap-filling methods (Section 5.2.3) to recreate observed surface ocean pCO_2 and that the mapping products based on multi-platform integrated systems yield results that best explain both zonal and meridional sampling strategies. The implications of adding high-resolution USV-Saildrone observations on the seasonal cycle of surface ocean pCO_2 are consistently different for summer and winter months as highlighted by SC anomalies (Figure 5.3). These SC anomalies are far larger in winter than in summer, resulting from the large advantage the SOCAT shipboard measurements already have in the summertime in most Southern Ocean regions (Figure 4.1). Impacts of USV-Saildrone circumpolar sampling on the seasonal cycle are more variable and one circumpolar deployment is not sufficient, though with improvements in most regions (Figure 5.3a-b). However, the impacts of adding the meridional sampling networks are greater and less variable for wintertime across all regions (Figure 5.3c).

Furthermore, while performing better in winter than SOCAT + zUSV[3], the SOCAT + mUSV[4] system further improved the RMSE score with a margin of 9.94% in the whole Southern Ocean, 6.28% in the SAZ, 7.37% in the PFZ, and 11.84% in the AZ (Figure 5.3b-c; Table 5.2). By this, I demonstrate that coordinated integration of multi-platform sampling systems (see Section 5.2.2), particularly with a meridional sampling strategy, in the future is a realistic way to resolve seasonal sampling gaps in the Southern Ocean. Thus, this significantly reduces winter biases and uncertainties that exist in pCO_2 reconstructions in the region.

5.4.3 The sensitivity to sampling strategies

So far, in light of what has already been shown (see Section 5.4.1), the positive anomalies that kept persisting in the Indian Ocean basin, especially in the AZ (see Figures 5.5 and C.2), is the main focus of the discussion here as I attempt to find an answer to "why?". Note the evaluation of the spread in mean bias errors between surface ocean pCO_2 from various integrated systems and Nature Run has revealed regional biases up to $\pm$20 µatm (Figure 5.5). These biases persist when zonally averaged and show the SOCAT-only and multi-platform integrated reconstructions differing in uncertainties by 0.43–0.82 µatm in the AZ compared to

1.06–1.79 µatm in the same region when meridionally averaged (Tables 4.4 and 5.1). This shows also that the pCO_2 reconstructions are sensitive to both zonal and meridional sampling strategies, which confirms that the scale sensitivity does not only matter at a local level like in the sub-domain study in Chapter 3 but also in the Southern Ocean as a whole. However, these pCO_2 anomalies in the Indian basin could stem from different mechanisms ranging from pCO_2 relationships with its drivers to weaknesses and strengths of the ensemble members.

Given the singularity of the regression step (FNN) in the ensemble approach, the weaknesses and strengths of its members are strongly correlated with the difference in the clustering step having three different methods. Thus, looking at Figures 4.11, 4.12 and 4.14, for example, the hypothesis of the weaknesses and strengths of ensemble members as the potential source of the positive mean bias errors between Nature Run surface ocean pCO_2 and reconstructions cannot be entirely neglected. Although the clustering step in the two-step reconstruction does control whether the FNN regression step has too much freedom for the training data, the FNN-derived relationships between input variables and surface ocean pCO_2 observations are non-linear. Therefore, preventing overfitting remains challenging as in any machine learning application, which can also result in positive mean bias errors. Moreover, persistently positive deviations in the Indian sector of the AZ (Figures 5.5 and C.2) likely result from spatial and seasonal inconsistency in the relationship between pCO_2 and the drivers such as the mixed layer depth (MLD), which might be more challenging to machine learning models to reproduce.

When investigating the distribution of total dissolved inorganic carbon (DIC) in the Southern Ocean using observations, Ishii et al. (1998) found that the total DIC content of surface seawater in the Indian sector of the seasonal ice zone showed large spatial variability. The BIOPERIANT12 ocean model (Section 2.1) is representative of processes of interest this thesis tries to resolve, but I do not know if the model output values are right or wrong. The model simulation likely captures relatively large spatial variability of surface ocean pCO_2 ranging from 335 to 388 µatm between 20°E and 60°E, which is very likely another cause of the positive pCO_2 anomalies persisting in the Indian Ocean basin around 60°S (Figures 5.5 and C.2). Furthermore, the surface ocean CO_2 in the Southern Ocean is known to be very variable spatially and temporally (Thomalla et al., 2011; Landschützer et al., 2015; Gruber et al., 2019), stemming from the drivers of pCO_2 in the region (Gregor et al., 2018). For example, the proxies of CO_2 such as the sea surface temperature (SST) and chlorophyll-*a* (Chl-*a*) vary strongly across short gradients in the Southern Ocean, resulting in concomitant changes in the surface ocean CO_2 (Resplandy et al., 2018; Monteiro et al., 2015) (Resplandy et al., 2014; Monteiro et al., 2015). Therefore, the positive pCO_2 anomalies might stem from the deficit of the total DIC in the marginal ice zone including the AZ for which biological activities play an important role (Ishii et al., 1998; Thomalla et al., 2011), which suggests that mechanism governing that may be the biological carbon pump.

In general, all the OSSE scenarios highlighted these positive mean bias errors around the boundary of the Atlantic and Indian basins of the Southern Ocean , which also seems to be particularly consistent in a few areas of the AZ (Figures 5.5 and C.2). The Atlantic sectors of the SAZ and PFZ have the strongest seasonal variability, contrasting the relatively weak seasonal amplitude of pCO_2 in the Indian sector of the Southern Ocean as reported by Gregor et al. (2018). According to the authors, the summer pCO_2 variability in the Indian basin spans a short period, providing a decrease in winter surface ocean pCO_2. Looking at the mean annual maps of reconstructed surface ocean pCO_2 in Figure C.1, this contrasting regional variability can also be spotted.

5.4.4 Overall comparison of OSSE sampling scenarios

In the following, I am discussing the performances in terms of both RMSE and MBE scores of the ensemble average learning in the five main OSSE sampling scenarios from both Chapter 4 (SOCAT-only, SOCAT + WGM) and this chapter (SOCAT + zUSV[1], SOCAT + zUSV[ALL] and SOCAT + mUSV[ALL]). The SOCAT-only scenario is the baseline or placebo experiment in this discussion (Figure 5.6).

The SOCAT + WGM sampling outperforms all the SOCAT + USV sampling experiments in the whole Southern Ocean and regionally (Figure 5.6; Tables 4.4 and 5.1) as expected. This is because the SOCAT + WGM sampling scenario is not only feasibly the most extreme case where I assumed that underway ships could sample along the same lines that the underway SOCAT ships use in summer but also because the ship in the WGM lines sample the ice-covered AZ in winter whereas the USV Saildrones do not (cf. Figures 4.1 and 5.2).

The difficulty of the feasibility of the SOCAT + WGM scenario in order to close the seasonal sampling gaps in the whole Southern Ocean is the main reason why USV Saildrones (resulting from Chapter 5) were brought into the picture to supplement SOCAT-only sampling, particularly using an orthogonal or meridional sampling strategy (mUSV). Thus, SOCAT + mUSV outperforms the current zonal circumpolar sampling strategy (SOCAT + zUSV) in the whole Southern Ocean as expected, but not in all the regions as shown in Figure 5.6. More specifically, an unexpected difference was found between SOCAT + mUSV and SOCAT + zUSV in the Antarctic zone (AZ) as shown in Figure 5.6a. Based on our understanding, this would be because practically, USV Saildrones hardly reach the southmost areas of the AZ as depicted in Figures 5.1b and 5.2. This reveals the impact of the AZ sampling gaps, for example, in the orthogonal sampling strategy of USV Saildrones.

This unexpected impact of the AZ sampling gaps shows how sensitive the RMSE and MBE scores are to sampling gaps in the Southern Ocean. Therefore, SOCAT + WGM consistently outperforms the SOCAT + USV samplings because of that, which reveals that USV Saildrones do not answer all the observational needs in the Southern Ocean. This finding suggests that winter underway ship sampling is necessary at least

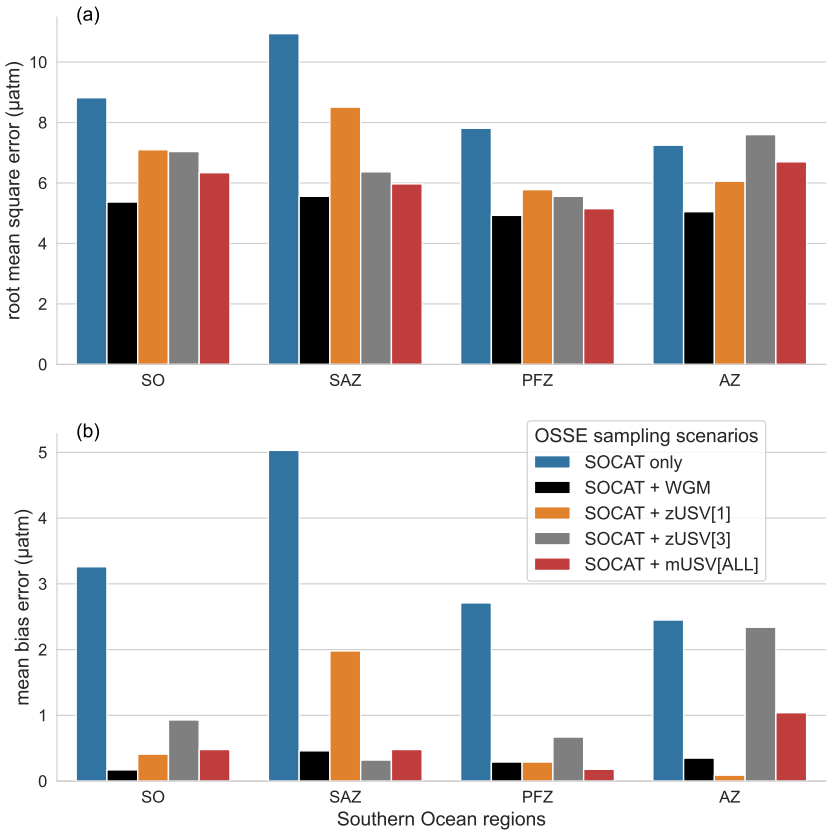


Figure 5.6: Comparing performances (y-axis) of the ensemble method in the five main OSSE sampling scenarios including SOCAT-only, SOCAT + WGM, SOCAT + zUSV[1], SOCAT + zUSV[3] and SOCAT + mUSV[ALL]. Panels **(a)** and **(b)** are respectively the root mean squared error (RMSE) and mean bias error (MBE) scores (y-axis). In the x-axis are the Southern Ocean regions, where the abbreviations are SO = Southern Ocean; SAZ = sub-Antarctic zone; PFZ = polar frontal zone; and AZ = Antarctic zone. It shows that all the OSSEs that included either a winter ship line or USV performed better that the SOCAT-only OSSE either in terms of RMSE scores or MBE scores.

in a couple of underway SOCAT ship lines that are operational in summer.

At this point, there is sufficient evidence that scale-sensitive sampling strategies which address limitations including winter and basin-scale sampling gaps as well as seasonal sampling aliases in high-EKE and mesoscale-intensive ocean regions, allow us to significantly reduce biases and uncertainties in mapped pCO_2. These findings reveal that the pCO_2 reconstruction in the Southern Ocean is very sensitive to sampling strategies that various observing platforms adopt in the region. It is thus time to "get over the wall".

5.4.5 Requirements to "get over the wall"

The "wall" metaphor identified by Gregor et al. (2019) is mainly due to the limitations of underway ship observations in the Southern Ocean. The analysis of the SOCAT-only product highlighted the major issues of these ship-based observations in the region stemming from the data scarcity in both space and time,

especially during the winter months (as shown in Figures 4.1, 4.14 and 5.3). This analysis extends the findings Bushinsky et al. (2019) and Gregor et al. (2019) who hypothesized that seasonal sampling gaps in the Southern Ocean might be increasing the pCO_2 mapping errors in the region. Despite these limitations, SOCAT data remains the best existing high-quality in-situ CO_2 observations that are fundamental to novel machine learning methods for global and regional surface ocean pCO_2 mapping.

Beyond increasing the data coverage

In Chapter 4, the machine-learning ensemble average of clustering-regression approaches was found to be more powerful than single methods and also tended to better mitigate under-/over-estimation in the reconstructions – based on the summer-biased SOCAT-only scenario (see Figure C.2). However, in zonally integrated systems such as SOCAT + zUSV[3] (see Figure 5.2b), USV Saildrones navigation is simpler because they are deployed following the wind direction and along the Antarctic Circumpolar Currents. This sampling was used in Sutton et al. (2021) (see Figure 5.1b). Therefore, increasing the spatial sampling and subsequently the data coverage would be easier than sampling along meridians where USVs will need to withstand Southern Ocean winds and waves (Meinig et al., 2019). But with just the zonal sampling strategies, the meridional gradients of pCO_2 are hardly resolved, as found in Chapter 3. Although they may hardly reach a large spatial extent (Lenton et al., 2006), USVs in meridional integrated systems such as SOCAT + mUSV[4] (see Figure 5.2c) can simultaneously observe at high frequencies of the seasonal cycles and meridional gradients of surface ocean pCO_2 as well as addressing the intra-seasonal variability (Djeutchouang et al., 2022).

Adopting a supervised orthogonal sampling for USV Saildrones

Results indicate that adding USV-Saildrone observations to the underway SOCAT ship observations improves the reconstructed pCO_2 in the Southern Ocean as a whole in terms of RMSE and MBE scores. This is consistent with the analysis in Chapters 3 and 4 and also extends our understanding of a previous assessment of the summer-biased observational impacts on air-sea CO_2 fluxes in the Southern Ocean by Bushinsky et al. (2019). Additionally, the smallest RMSE scores were found for the SOACT + mUSV[4] product (cf. Table 5.1), revealing that the meridional sampling strategy contributes best to reducing uncertainties in the mapped pCO_2 in the Southern Ocean as a whole. On the other hand, the positive implications of this orthogonal sampling strategy on the mean seasonal cycles of surface ocean pCO_2 were significant (cf. Figure 5.3c). This positive influence was particularly remarkable in data-sparse areas such as the Pacific sector of the SAZ where the zonal circumpolar strategy underperformed (cf. Figures 5.3 and 5.5; Tables 5.1 and 5.2) despite having the largest number of data points. In the AZ, however, both zonal-circumpolar and meridional sampling strategies of USV Saildrones can be adopted for scale-sensitive samplings as in both cases USV Saildrones struggle to reach the southmost areas of this region (Figures 5.1 and 5.2) due to the seasonal ice zone. This demonstrates Chapter 3's conjecture for which simultaneously samples at high resolution the

meridional gradients and the seasonal cycle of pCO_2^{ocn}, achievable by USV Saildrones, improves biases and uncertainties in pCO_2 reconstructions in the Southern Ocean as a whole.

Therefore, "getting over the wall" in terms of reducing biases and uncertainties in observation-based pCO_2 reconstructions requires the community to start implementing supervised meridional sampling strategies with autonomous sampling platforms such as USV Saildrones. However, this must not discontinue the underway SOCAT ship observations in the Southern Ocean. For example, in areas unreachable by the ships and seasons where they cannot operate, USV Saildrones can deliver these meridional sampling routes.

5.5 Conclusions

In this chapter, I propose a new supervised sampling strategy that significantly reduces the uncertainties and biases of empirical reconstructions of the surface ocean pCO_2 in the Southern Ocean, therefore, "getting over the wall" remarked by Gregor et al. in their 2019 article. This is done with a multi-platform integrated sampling of the surface ocean which improves training datasets – used by gap-filling methods – by combining pseudo-observations from underway SOCAT ships and deployments of high-resolution USV Saildrones in the Southern Ocean. The feasibility was demonstrated with an OSSE approach whose Nature Run was the BIOPERIANT12 ocean model outputs like in Chapters 3 and 4. The experimental combination of training datasets was made off of two USV-Saildrone sampling strategies including a zonal circumpolar sampling and a meridional sampling of the Southern Ocean. Addressing and resolving this issue required us to answer the following questions raised in Section 5.1.1:

1. **Do simultaneous observations of the seasonal cycle of the meridional gradients of pCO_2^{ocn}, which USV Saildrones are capable of providing, significantly improve the biases and uncertainties in pCO_2 reconstructions in the Southern Ocean as a whole?**

The underway SOCAT ship data was integrated with USV-Saildrone data for simultaneously resolving both high-resolution observations of the meridional gradients and the seasonal cycle in the Southern Ocean. It was found that this improved training dataset contributed to a significant reduction of pCO_2 mapping biases and errors in the whole Southern Ocean and regionally. More explicitly, this led to a 20% and 28% decrease in pCO_2 reconstruction errors in the whole Southern Ocean when using zonal circumpolar and meridional sampling strategies, respectively. Regardless of the sampling strategy adopted by the USV Saildrones in the Southern Ocean, a year-round deployment can provide observations of the full seasonal cycle (SC) in the region. When combining these high-resolution observations with underway SOCAT ship measurements, the findings show improvements during the winter months in the reconstructed mean seasonal cycle of surface ocean pCO_2 throughout the Southern Ocean.

2. **Would meridional sampling with USV Saildrones in the Southern Ocean (as a whole) make any**

differences as compared to the currently used zonal circumpolar strategy?

The two sampling strategies adopted by the USV Saildrones in the Southern Ocean – the current zonal circumpolar sampling and a newly proposed meridional sampling – lead to reducing uncertainties in pCO_2 estimates in the region. However, when comparing both strategies, results of the analysis reveal that the meridional sampling outperforms the current zonal circumpolar sampling in the Southern Ocean as a whole. In addition, with this sampling strategy, significant differences are made in terms of reducing winter anomalies in the reconstructed seasonal cycle (SC) throughout the Southern Ocean regions. These findings suggest that in the Southern Ocean, the sampling scale sensitivity is the major problem rather than the quantity of data.

With these scale-sensitive sampling strategies, the pCO_2 SC anomalies due to winter sampling gaps in the Southern Ocean are significantly reduced across regions. However, these improvements of SC anomalies during wintertime in the SAZ are not as much for the current zonal circumpolar sampling strategy as for the meridional sampling. Although the strongest errors occur during winter, increasing sampling effort in winter is not sufficient to improve pCO_2 estimates and the mean seasonal cycle throughout the Southern Ocean, particularly in the Pacific sector of the SAZ. To improve reconstructions of surface ocean pCO_2 in the Southern Ocean as a whole, sampling scales of current and future observing platforms must be improved. A multi-platform integrated system combining simultaneous high-resolution observations of the meridional gradients and the seasonal cycle of pCO_2^{ocn} – achievable by USV Saildrones for example – and underway SOCAT ship data is proposed here as a realistic sampling strategy to explore in the future in order to improve pCO_2^{ocn} mapping in the Southern Ocean. However, it was found that USV Saildrones cannot answer all the observational needs in the Southern Ocean, particularly in the AZ which is a source of sampling gaps for USV Saildrones because of the seasonal ice zone. This suggests that winter underway ship sampling is necessary at least in a couple of the underway SOCAT ship lines that are operational in summer.

In summary, leveraging the strengths of autonomous ocean observing platforms such as USV Saildrones (high speed, wind- and solar-powered, etc.), high resolutions of year-round meridional observations can be maintained in high-EKE regions of the Southern Ocean and most importantly in areas where contemporary sampling approaches – based on underway SOCAT ship routes – cannot be conducted. This vertical resolution sampling strategy with autonomous observing platforms has not yet been tested and the ocean carbon community in the Southern Ocean could be tasked to explore this to supplement ship-based hydrographic surveys.

Chapter 6

Synthesis, limitations and recommendations

In this study, I investigated the sampling scale sensitivities of surface ocean pCO_2 reconstructions in the Southern Ocean derived from machine-learning techniques. It was done through a series of observing system simulation experiments (OSSEs) applied to a forced mesoscale resolving ($\pm$10km) ocean NEMO-PISCES physics-biogeochemistry model (BIOPERIANT12) with daily output introduced in Section 2.1. The analysis of the multiple experiments indicates that improving the pCO_2 reconstructions requires scale-sensitive data to to strengthen both the existing underway SOCAT-ship and SOCCOM-float observational products. It was also found that scale-sensitive data consisting of high-resolution observations ($<$ 1 days) extending over the seasonal cycle and capturing the pCO_2 meridional gradients – resulting in "getting over the wall" (Gregor et al., 2019).

6.1 Synthesis of findings and implications

The overarching aim was to investigate the sampling scale sensitivity of surface ocean pCO_2 machine-learning reconstructions in the Southern Ocean, a spatially and temporally data-sparse but globally important region for ocean carbon sink (Gregor et al., 2019; Gruber et al., 2019). This aim has been difficult to achieve with conventional in-situ ship-based pCO_2 observations from the SOCAT database due to their summer seasonal sampling biases (Bushinsky et al., 2019; Gregor et al., 2019; Gray et al., 2018). As the main objective, I have addressed an important question for the global carbon cycle community: how to reduce uncertainties and biases in contemporary observation-based reconstructions of surface ocean pCO_2 in the Southern Ocean derived from machine learning techniques. Machine learning methods have "hit the wall" in terms of improving pCO_2 reconstruction uncertainties and biases (Gregor et al., 2019). I found two possible reasons explaining why these methods are "hitting the wall". This includes the lack of (1) observations for the seasonal cycle of the pCO_2 meridional gradients and (2) the intra-seasonal variability of pCO_2.

I used a series of OSSE scenarios to investigate this question of the scale-sensitivity of pCO_2 reconstruction uncertainties. The chosen OSSE Nature Run (Section 4.2.3) is assumed representative of the carbon cycle processes in the upper Southern Ocean. It thus represents the scales of variability of surface pCO_2 aimed to be examined in the thesis and does not represent any specific historical year. Using a framework of OSSEs, I made a proposal on how to get over the "wall" by closing the seasonal sampling gaps in the Southern Ocean through a viable hybrid sampling strategy that strengthens the underway ship-based SO-

CAT data with winter meridional lines by well-placed autonomous unmanned surface vehicles (USVs) in the Southern Ocean. This hybrid sampling strategy require scales-sensitive sampling in the Southern Ocean; that is, sampling at as high-resolution as necessary.

6.1.1 Does scale-sensitive sampling matter in pCO_2 reconstructions?

Sparse sampling in both space and time in the Southern Ocean has raised questions about the confidence in present machine-learning reconstructions of pCO_2 derived from seasonally biased sparse observations in the region (Bushinsky et al., 2019; Gregor et al., 2019; Gloege et al., 2021), but previous studies did not yield a clear understanding of the uncertainties and biases. This question was examined in Chapter 3 through a series of semi-idealized OSSEs, within the Nature Run sub-domain of 10° latitude and 20° longitude including both the sub-Antarctic zone (SAZ) and polar frontal zone (PFZ) in the south-east Atlantic Ocean, which are the two most sampled sub-regions of the Southern Ocean. In these OSSEs, observational scales of surface ocean pCO_2 were simulated in ways that are comparable to existing ocean CO_2 observing platforms including ships (SHIP), Wave Gliders (WG), carbon floats (FLOAT) and Saildrones (nUSV). It was shown that while this sub-domain is small relative to the Southern Ocean scales, it was representative of the scales of variability of interest including strong mesoscale gradients interacting with synoptic scale variability of storms.

In these experiments, it was found that two sampling scales are required to improve pCO_2 reconstructions using machine-learning techniques. The two sampling scales are (1) the seasonal cycle of the meridional gradients and (2) the intra-seasonal variability. Based on the impacts of these two sampling scales on RMSE scores, resolving the seasonal cycle of the meridional gradient is the first-order requirement, while resolving the intra-seasonal variability is the second. However, it is important to note that additional timescales (beyond the seasonal cycle) cannot be resolved because the mesoscale-resolving ocean model run used in the OSSEs is limitated to only one year. For SOCAT-ship pseudo-observations, other studies including Gloege et al. (2021) have addressed inter-annual and decadal timescales by sub-sampling a few medium-resolution Earth system models that do not resolve the sub-grid scale dynamics which are explicitly simulated in the BIOPERIANT12, the mesoscale-resolving ocean model used in the thesis.

In this Nature Run sub-domain, improving pCO_2 reconstructions in the Southern Ocean required simultaneous high-resolution observations (< 1 days) of the whole seasonal cycle, with full resolution of the meridional gradients of pCO_2. Saildrones were an optimal sampling platform to address both the large-scale spatial and high-resolution temporal sampling and had the most effective impact on reducing uncertainties and biases of the seasonal and annual mean reconstructions of pCO_2. It was also found that the addition of wintertime underway ship data would greatly reduce the errors in the reconstructions of surface ocean pCO_2, which reveals that the present seasonal sampling biases in the SOCAT data in the Southern Ocean

may be behind the biases in wintertime that observation-based data products highlight in this region.

Therefore, based of these findings, scale-sensitive observations have significant impacts on surface ocean pCO_2 reconstructions. Chapters 4 and 5 focused on the first-order requirement which consists of observing the seasonal cycle of the meridional gradients of surface ocean pCO_2 in the Southern Ocean. Specifically, Chapter 5 also proposed a scale-sensitive sampling strategy for the whole Southern Ocean by integrating Saildrones with underway ships on winter lines.

6.1.2 Clustering choice in the Southern Ocean: sensitivity to two-step pCO_2 reconstruction

The large zonal spatial gaps with high dynamic variability between the meridional lines in the Southern Ocean are a significant challenge, which is addressed by the initial clustering step of the two-step reconstruction approach used in this study (Landschützer et al., 2014; Gregor et al., 2019). Answering the scale-sensitive sampling question for the Southern Ocean as a whole also required us to run OSSEs in the whole Southern Ocean to evaluate clustering choices in two-step pCO_2 reconstructions through two OSSE summer and winter sampling scenarios. This provided the opportunity to also explore the impacts of winter observational gaps on reconstruction outcomes.

Clustering-regression approaches are the state-of-the-art for observation-based pCO_2 estimates in various ocean carbon and biogeochemistry Working Groups in the global carbon community (Landschützer et al., 2016; Gregor et al., 2019; Denvil-Sommer et al., 2019; Gregor and Gruber, 2021). However, in filling the gaps in contemporary observation-based estimates of air-sea carbon fluxes with these two-step methods, the lack of out-of-sample data is a serious challenge, particularly in data-poor regions like the Pacific Ocean sector. The clustering choice must thus be made with diligence. The use of OSSEs to evaluate clustering methods, the first step in robust two-step reconstructions of surface ocean pCO_2, proved necessary in Chapter 4. This was achieved by implementing three clustering-FNN regression methods from (1) Fay and McKinley (2014) biomes (FMB), (2) mini-batch K-Means (MKM) and (2) seasonal cycle reproducibility (SCR) of pCO_2. Two OSSE sampling scenarios were considered: (1) SOCAT-only, which only simulated the summer-biased underway SOCAT ship observations from the Nature Run; and (2) SOCAT + WGM, where an idealized ship-based sampling strategy called winter gap mitigation (WGM) was added to SOCAT-only to compensate for the winter observational gaps.

It was shown that while these cluster-learning approaches perform comparably throughout the Southern Ocean, the ensemble average learning (ENS) combining the three approaches together significantly improves the results and makes findings more robust and less clustering-method specific. The use of SCR clustering in this thesis is not only a novelty but is also the only method out of the three clustering methods

that explicitly considers the spatial characteristics of the seasonal cycle of surface ocean pCO_2. This is very important knowing that seasonality and its intra-seasonal modes can dominate variability in the Southern Ocean. Therefore, the use of the physical-biogeochemical biomes (Fay and McKinley, 2014) and seasonal cycle reproducibility sub-regions in the neural network represented a mean that improved on either one individually. The additional improvement from adding mini-batch K-Means derived clusters likely adds extra information to the reconstructions that FMB and SCR may be lacking and reversibly. This highlights the careful consideration of questions of spatial and temporal scale, especially where their interaction enhances the mixed layer dynamics that influence the pCO_2 gradient ΔpCO_2 and the CO_2 flux F_{CO_2}.

Chapter 4 showed that an ensemble of three clustering methods, the first step in the two-step pCO_2 reconstructions, outperformed each individual method to reconstruct pCO_2 in the Southern Ocean. This highlighted that the clustering ensemble could also contribute towards significantly reducing the RMSE and MBE scores. This points to the need of a scale-sensitive sampling strategy that addresses the limitations and strengths of the existing summer-biased ship-based and observation-based reconstructions in order "to get over the wall". The sensitivity of pCO_2 reconstructions was tested for two sampling strategies using USV Saildrones in the Southern Ocean zonal circumpolar sampling and new orthogonal sampling strategies. An OSSE framework was used to address the resulting questions (Section 5.1.1). It was found that the contribution of the clustering methods to uncertainties of pCO_2 reconstructions showed a seasonal sampling bias strength. For example, the percentage decrease in the ENS RMSE scores from the SOCAT-only scenario to SOCAT + WGM was around 30% in general for most of the regions, up to 47% for the sub-Antarctic zone (SAZ). On the other hand, the anomalies of mean seasonal cycle estimates in the Southern Ocean revealed that if data are added in winter (e.g.: SOCAT + WGM), then the biases would drop down by about 85%.

Overall, these findings propose that an ensemble approach to clustering methods in two-step observation-based reconstructions of surface ocean pCO_2 in the Southern Ocean can improve estimates beyond the bounds of the current observations.

6.1.3 Getting "over the wall" through sampling of the meridional gradient in both summer and winter

Chapter 3 highlighted sampling scales as the source of uncertainties and biases in pCO_2 reconstructions in the Southern Ocean. More specifically, it highlighted two sampling scales of variability including the seasonal cycle of the meridional gradient and the intra-seasonal variability which are critical to reducing uncertainties and biases in pCO_2 reconstructions in the region. It showed that the seasonal cycle of the meridional gradient was the first-order requirement and the intra-seasonal was second-order. This resulted from findings within a small experimental model sub-domain of the Southern Ocean that high-resolution orthogonal sampling with USV Saildrones was the most optimal and realistic strategy.

It was found that a supervised integration of underway ships with high-frequency USV Saildrones simultaneously resolves high-resolution observations of the seasonal cycle of meridional gradients of surface ocean pCO_2 in the Southern Ocean which contributes to a significant reduction of pCO_2 reconstruction biases and uncertainties in the whole Southern Ocean and regionally. It was also found that the meridional sampling with USV Saildrones outperforms the current zonal circumpolar sampling in the Southern Ocean as a whole with a margin of about 10% in the percentage change of their RMSE scores.

Moreover, with recent technological advances in autonomous devices, the results of this thesis can help to guide the design of new physical and biological measurement sensors and instruments, and ultimately to determine if a new CO_2 observing platform will be cost-effective. There have been tremendous efforts to improve sampling coverage in the Southern Ocean, but this thesis has shown that what matters the most is not the quantity of surface ocean CO_2 observations but the sampling scales in both the high frequency and meridional extent of these observations. This, I think, will help us get over this "wall" in terms of improving pCO_2 reconstruction biases and uncertainties but will no doubt reveal further walls.

6.1.4 Proposed sampling strategy for the Southern Ocean

Certain regions of the Southern Ocean are characterized by seasonal variability and will not require to be sampled at high temporal resolution such as better than 1 day. In these regions, the 10-day observational frequency typical of floats (FLOATs), for example, would be sufficient. Here, autonomous observing platforms such as the current existing floats can be utilized in these regions characterized by seasonal variability. In Chapter 5, the results show that if FLOATs sampling frequency is increased to 1 - 3 days that their impact on reducing RMSE scores can be significantly improved by about 73% and 79% when deploying them in the SAZ and PFZ, respectively.

However, in regions characterized by spatial variability, the commonly used 1°-by-1° gridding averages pCO_2 values over those scales and that is a problem where we have strong frontal mesoscale gradients such as the sub-Antarctic front. Therefore, a scale-sensitive sampling is required to close the spatially and temporally observational gaps in the Southern Ocean. Based on the findings, I proposed a hybrid sampling strategy that should reduce the RMSE and MBE to 30% by integrating summer underway ship sampling with USV Saildrones for winter lines and 1-day period floats in regions of high dynamic variability and high EKE activities.

6.2 Limitations of the study

USV Saildrones stand out to be essential because currently, it is the optimal autonomous ocean observing platform whose strengths (high speed, wind- and solar-powered, etc.) can be leveraged to get the gradients

of surface ocean pCO_2 as a substitute of the underway sampling along cruise tracks. The problem with other autonomous devices such as FLOATs is that they follow the general circulation, and the Lagrangian currents are zonal. In Chapter 3, it was found that this zonal sampling strategy was unable to resovle the meridional gradients of surface ocean pCO_2. However, even with USV Saildrones, there are not yet practical and sustainable ways to pilot USV Saildrones against strong winds and across intense fronts in the Southern Ocean. Therefore, when simulating orthogonal sampling from these platforms in the Southern Ocean (Chapters 3 and 5), the main assumption was that they were capable to resist the strong westerly winds. This assumption, though not impossible to satisfy, constitutes one of the main limitations of our study because, in the real world, winds are the major barrier to meridional sampling in the Southern Ocean, which only research vessels have been able to carry out successfully.

It is not straightforward to determine the exact influence of underway SOCAT ship observations versus the USV-Saildrone observations, as the FNN-derived relationships between input variables and surface ocean pCO_2 observations are non-linear. Therefore, these relationships hardly improve based on the absolute observation size. Moreover, in data-sparse regions such as the Pacific Ocean sector of the Southern Ocean, complex machine-learning algorithms are prone to overfitting the surface ocean pCO_2 data. In such regions, using less complex regression algorithms such as multi-linear regression might reduce the risk of overfitting.

The forced ocean model used as Nature Run in the OSSE sampling scenarios is of sufficient spatial resolution to capture the main dynamics of the Southern Ocean, but it might not yet represent correctly the spatial variability of surface ocean pCO_2 in the Southern Ocean. Due to this limitation, capturing the impact of sub-mesoscale variability in regions of high dynamic variability remains uncertain in spite of the model output being representative of the processes of interest in this thesis. For example, the BIOPERIANT12 ocean model sub-sampled does not capture the seasonal cycle of pCO_2 suggested by observation-based reconstructions as presented in Figure 3.2. Given the aim of the thesis which is about investigating the sensitivity of reconstructed mean annual of pCO_2 to sampling aliases (see Section 1.1.2), it was assumed that it does matter if the model appears incorrect in terms of capturing the phasing of the seasonal cycle of pCO_2 data products (Gregor et al., 2019; Landschützer et al., 2016). Even though the BIOPERIANT12 model represents scales variability of interest, this is a limitation which might have some implications on such experiments if the goal was to reconstrunct the seasonal cycle. Thus, if the seasonal cycle is not captured, it could mean that the mesoscale would be driving away from the expected climatology. Furthermore, due to the high dynamic variability and high eddy activities of the Southern Ocean, simulations in the OSSEs are less likely to correctly match the spatial scales of underlying observing platforms. Therefore, simulated observational scales of surface ocean pCO_2, although comparable to underlying ocean CO_2 observing platforms in terms of their temporal sampling scales, are likely biased towards their spatial ones.

Moreover, in the Southern Ocean, which is known for its strong and persistent westerly winds, Rossby's deformation radius is particularly relevant for understanding the behavior of eddies. The strong westerly winds create a gradient in the ocean's rotation, which leads to the deformation of eddies with scales that are on the order of Rossby's deformation radius (Taylor et al., 2018). These eddies, known as mesoscale eddies, play an important role in the dynamics of the Southern Ocean, as they transport heat and carbon across the ocean (Taylor et al., 2018; Frölicher et al., 2015; Sallée et al., 2013; DeVries et al., 2017). Given that the size of eddies is set by Rossby's deformation radius which is smaller than 10 km in the Southern Ocean, eddies in the region will likely be smaller. Therefore, whether 10 km is sufficient for the BIOPERIANT12 ocean model to be an eddy-resolving ocean model or not becomes a possible area of the study's limitations.

Despite the random sub-sampling of the model outputs, the resulting pseudo-observations might not be independent if two neighboring measurements end up in the training and validation datasets. For example, Jones et al. (2012) show that the autocorrelation length of ship-based measurements is 400 ± 250 km. Therefore, the spatial autocorrelation can temper the independent assumption of training and validation samples. In case, if this assumption is violated, some training and validation samples would basically be as if we used the same data due to their proximity in space. This issue is independent of using best practices in the training of machine learning models and has likely affected other experiments that used SOCAT for the reconstruction of surface ocean pCO_2. This is also one of the problems the clustering step, in the two-step reconstruction approach, tries to resolve. However, with multiple-year data, the split between training and validation subsets based on a random subset of years in the time series instead of randomly sub-sample the whole data will resolve a similar issue which could emerge due to auto-correlation in time.

6.3 Recommendations for future work

This thesis addressed a potentially important gap in our current knowledge of the ocean carbon sink estimates and provided suggestions on how the global carbon cycle community can improve current estimates of the Southern Ocean carbon fluxes, through an improved sampling strategy and clustering approach.

6.3.1 Two-step reconstructions and ensemble clustering

The supervised clustering approach, based on the seasonal cycle reproducibility of surface ocean pCO_2 in the Southern Ocean, proposed in Chapter 4 needs to be tested globally. This will require high-resolution physics-biogeochemical ocean model outputs for the OSSE synthetic truth in order to resolve intra-seasonal variability. As shown in Chapter 3, intra-seasonality, if not resolved, constitutes a challenging factor in reducing uncertainties and biases of the mean seasonal cycle of CO_2 fluxes estimates due to temporal sampling aliases.

One general caveat is that the numerical ocean biogeochemistry model providing sufficiently high-resolution output is suitable for representing the environmental reference or synthetic truth in the OSSE. This allows the sub-sampling of the training datasets in various OSSE sampling scenarios and withholds enough datasets to effectively test all the underlying methods. Given a clustering choice, the first step in the two-step pCO_2 reconstructions controls whether the FNN regression step has too much freedom for the training data. A substantial analysis in Chapter 4 revealed that an ensemble of clusters, rather than a single one, was to have the greatest impact on the reconstruction uncertainties and biases. Based on these results, I recommend an ensemble of clustering methods in the two-step reconstruction of pCO_2 in the Southern Ocean.

6.3.2 Scale-sensitive sampling: closing the seasonal and intra-seasonal gaps

Unlike other autonomous sampling platforms such as floats, USV Saildrone deployments can be built from prior data-coverage knowledge in a data-sparse region. However, though not necessary, USV-Saildrone planned tracks can be based on prior knowledge of water mass properties, mean frontal locations (zonal sampling), mean circulation and eddy variability (both zonal and meridional sampling), winds (meridional sampling), air-sea heat/carbon exchange, and USV simulations in the Southern Ocean. In the development of our understanding of the importance of scale-sensitive sampling in reducing biases and uncertainties of pCO_2 reconstructions in the whole Southern Ocean and regionally in Chapters 3, 4 and 5, it was shown that USV-Saildrone observation at high-resolution of the seasonal cycle of the meridional gradients of surface ocean pCO_2 was necessary. However, given that these experimental results are all based on synthetic truth, I recommend the application of the OSSE framework with different numerical ocean biogeochemistry models to validate the findings.

However, when comparing the overall OSSE sampling scenarios in Section 5.4.4, results particularly highlighted the impact of the Antarctic zone (AZ) sampling gaps in the various sampling strategies of USV Saildrones due to their inability to reach the southmost areas of AZ. This was the reason why SOCAT + WGM consistently outperforms the USVs, which reveals that USV Saildrones do not answer all the observational needs in the Southern Ocean. This proves that in closing the seasonal sampling gaps in the Southern Ocean, winter underway ship sampling in the sea-ice zone is still necessary, but it is not clear which sector (Pacific, Atlantic and Indian) and how many meridional winter lines are necessary to supplement the proposed USV Saildrone lines. This is an issue that requires further investigation.

Resolving intra-seasonal scales in Chapter 3 resulted from the high sampling frequency of USV Saildrones and Wave Gliders. Argo floats equipped with biogeochemical sensors, require more coordinated deployments to resolve the intra-seasonal variability than USV Saildrones and Wave Gliders can resolve due to their hourly sampling frequency. Therefore, sampling the surface ocean pCO_2 at the correct interval is critical. That is why, based on these results, I recommend a 1-2 day sampling period for autonomous

observing platforms to complement the underway ship sampling and close the seasonal sampling gaps in the Southern Ocean. Moreover, based on the sampling period sensitivity analysis, Monteiro et al. (2015) showed that to achieve a 10% uncertainty threshold in the Southern Ocean as indicated by Lenton et al. (2006) as necessary, a sampling period of 1-3 days was required for areas of high eddy kinetic energy and elevated sub-seasonal dynamics. It remains an open research question how to close the seasonal sampling bias in surface ocean CO_2 observations in the Southern Ocean. Therefore, what would happen if winter underway ship sampling occurred on routes different from underway SOCAT ship routes requires further investigation.

Furthermore, given the focus of the thesis as stated in Section 1.1.2, what is important is that the BIOPERIANT12 ocean model is correct in terms of modes of variability and not necessarily in absolute magnitudes. The BIOPERIANT12 ocean model reflects the characteristics of variability intra-seasonal vs seasonal, observed from ocean-observing robotics (Monteiro et al., 2015; Nicholson et al., 2022). Therefore, a question requiring further investigation would be whether those modes of variability are present in the four selected Earth system models (ESMs) used in Gloege et al. (2021). A weaker simulation of intra-seasonal variability in the ESMs would translate into models performing well at the seasonal cycle mode. Notwithstanding the evidence provided by Gloege et al. (2021) regarding the robustness of pCO_2 reconstructions at the seasonal cycle mode, it does not matter whether the model appears to be correct because either way, the machine learning algorithms used in the reconstructions will try to capture the hidden features (such as the seasonal cycle) of the data it has been trained on.

6.3.3 Observing system simulation and real-world application

There is a few considerations of the real-world application of the suggested observational network strategy due to the consideration of the observational limitation of satellites. When applying the model-world results into the real world, two important consideratios should be made: (1) satellite chlorophyll-*a* concentrations are not fully available in high-latitude regions during winter, and (2) satellite observations are not commonly available as less than 3-day merged products. Therefore, I think taking Chlorophyll-*a* climatology is the best possible option to achieve the reconstruction of temporally complete fields as it is usually done for the mixed layer depth (Landschützer et al., 2014). This approach is being used in major observational-based reconstructions of surface ocean pCO_2 using machine learning techniques (Landschützer et al., 2016; Gregor et al., 2019; Gloege et al., 2021). However, this remains a serious caveat that I recommend further investigation in a future study. Moreover, while Gloege et al. (2021) focused on quantifying errors in observational-based pCO_2 reconstructions by sub-sampling the model using only SOCAT ship tracks, the thesis investigated not only the limitations due to seasonal sampling biases that ship-based observations present but also demonstrated how autonomous observing platforms could be leveraged in the future to sup-

plement ship-based observations. This will help reduce biases and uncertainties in pCO_2 reconstructions in the Southern Ocean. However, the BIOPERIANT12 ocean model used for this study does not captures current observation-based estimates of the season cycle phasing (cf. Section 3.2.3).

Therefore, it remains unclear how this limitation impacts the results. The impact of the reconstruction uncertainties and biases on the seasonal cycle on the interannual and decadal variability and trends, especially divergences in respect of forced ocean models (Hauck et al., 2020), needs to be investigated by the community. This investigation is a priority beyond this study and require multi-year runs of mesoscale-resolving ocean model simulations. Gloege et al. (2021) used medium-resolution Earth system models that do not resolve the sub-grid scale dynamics which are explicitly simulated in BIOPERIANT12, the ocean model used in the thesis. However, even if the medium-resolution models can improve reconstruction (Gloege et al., 2021), we will still need a high-resolution model like BIOPERIANT12 because it resolves the mesoscale which is critical to investigate the sensitivity of reconstructed mean annual of pCO_2 to sampling aliases. In addition, the authors looked at the performance against SOCAT data and did not take the same sensitivity analysis that this study does by including not only SOCAT ships but also autonomous sampling platforms. Thus, it is hard to draw direct critical evaluations of one approach against the other. A further study will be required to make an adequate comparative analysis on different timescales with Gloege et al. (2021)'s results, when multi-year runs of mesoscale-resolving ocean model simulations will be available.

The Southern Ocean is a very complex system with high dynamic variability and eddy activities. The complexity of the region's physical and biogeochemical systems involved in the ocean CO_2 sink makes it difficult to understand, assess and quantify. Independent observation-based estimates of these oceanic CO_2 sink magnitudes and their geographical distributions are significantly inconsistent and have large uncertainties and biases, particularly in the Southern Ocean, a data-sparse but globally important region. Therefore, if future greenhouse warming due to anthropogenic CO_2 emissions and associated climate changes are to be predicted, these observation-based estimates of CO_2 sinks in the ocean must be derived with relatively small errors, which also required improvements in the ways observations are done.

As our understanding of ocean systems and climate links improves with our increasing modelling capabilities and computational power, data collection has become more than ever perhaps the most important contribution of the global carbon community. The continuation and expansion of observational efforts in the Southern Ocean is a necessity due to its complex evolving nature and the vital role it plays in the non-stationary context of climate change through anthropogenic CO_2 uptake. Therefore, the emerging opportunity to use USV Saildrones for orthogonal sampling in the Southern Ocean areas and seasons where underway ships cannot reach needs to be further explored, particularly since these autonomous platforms provide, for the first time, hourly surface ocean pCO_2 observations from the circumnavigation of Antarctica.

Supplementary Assets

Asset A

Parameterization of the FNN and GBM regression

The Feed-forward Neural Network (FNN) is a class of neural network algorithms that is the most commonly used as anon-linear approach in the surface ocean pCO_2 reconstruction community (Bushinsky et al., 2019; Denvil-Sommer et al., 2019; Gloege et al., 2021; Gregor et al., 2019; Gregor and Gruber, 2021; Landschützer et al., 2016; Rödenbeck et al., 2015; Zeng et al., 2017). The Gradient Boosting Machines (GBM) is a widely used machine learning algorithm due to its efficiency, accuracy, and interpretability (Natekin and Knoll, 2013; Gregor et al., 2019; Gregor and Gruber, 2021; Ke et al., 2017). An implementation of GBM, called Light GBM was used in the thesis. Both machine learning algorithms (FNN and LightGBM) come with many hyper-parameters that are simply parameters whose values are not determined by the training but needed to control the learning process or training of an algorithm. A few of these hyper-parameters for both FNN and LightGBM are present in Table A.1.

Table A.1: A few hyper-parameters of both FNN and LightGBM algorithms and their usage.

Algorithms	Hyper-parameters	Usage/Descriptions
FNN	*loss function*	Determines how "quickly" the gradient updates follow the gradient direction.
	learning rate	Compares the network's output for a training example against the intended ground truth output.
	number of hidden layers	Large hidden layers cqwertyan allow the neural network to fit the training data arbitrarily well, but because regularization is typically used, it is mostly important to just use large hidden layers.
	number of training iterations	Used in the "early stopping" principle to simply stops training once performance on a held-out validation set stops increasing.
LightGBM	*maximum depth*	Sets a limit on the depth of tree. It is effective in controlling overfitting.
	number of leaves	Specifies the number of leaves in a tree
	bagging fraction	Specifies the fraction of data to be considered for each iteration.
	number of iterations	Specifies the number of iterations to be performed.
	maximum bin	Specifies the maximum number of bins to bucket the feature values.
	minimum data in bin	Specifies the minimum amount of data in one bin.
	categorical feature	Specifies the categorical feature used for training.
	feature fraction	Specifies the fraction of features to be considered in each iteration.
	task	Specifies the task we wish to perform which is either train or prediction.

Asset B

The SHIP, FLOAT, WG, nUSV Saildrone experiments

Table B.1: Summary of all the 8 semi-idealized ocean system simulation experiments (OSSE-8) conducted in this study. The simulated ocean observing platforms (SHIP, FLOAT, WG, and nUSV Saildrone) correspond to their real-world counterparts (ship, carbon-float, Waveglider, and Saildrone) used in the SOCAT project, SOCCOM initiative, and SOCCO program and by Saildrone Inc., respectively. The sampling regimes or scenarios represent the periods in which the data sampling phase of different experiments occurred according to the temporal scales of the underlying platforms. Note that the observing platforms Waveglider and float have two scenarios each based on the fact that they are deployed either in the north (SAZ) or south (PFZ) of the 10°-by-20° experimental domain. Experiment abbreviations together with their subsequent scenarios (defined by the sampling regimes/scenarios) are used in figures and throughout Chapter 3.

Ocean Observing Platforms	Sets	Sampling Regimes	Experiments
Ships (SOCAT-like)	SHIP	Summer (smr)	SHIP(smr)
		Summer + Winter (smr+wtr)	SHIP(smr+wtr)
		Autumn + Spring (aut+spr)	SHIP(aut+spr)
Floats (SOCCOM-like)	SHIP + FLOAT	Summer (smr) + One year round	SHIP(smr) + FLOAT(SAZ)
			SHIP(smr) + FLOAT(PFZ)
			SHIP(smr) + FLOAT(SAZ+PFZ)
Wavegliders (SOCCO-like)	SHIP + WG		SHIP(smr) + WG(SAZ)
			SHIP(smr) + WG(PFZ)
Saildrones (Saildrone Inc)	SHIP + nUSV		SHIP(smr) + nUSV

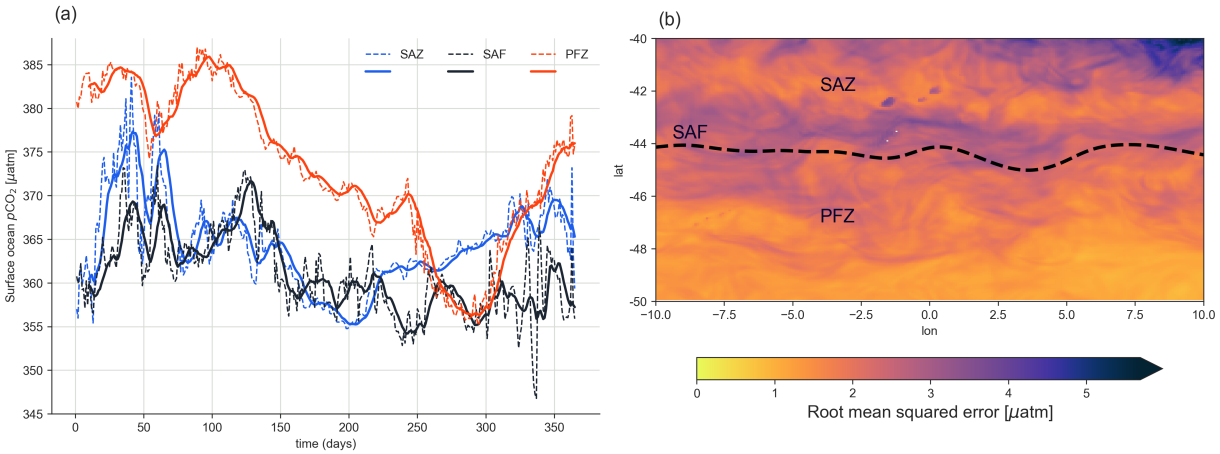


Figure B.1: **(a)** shows one-year time series plots (dashed lines) of the variability of pCO_2 at single-model grid cells on the SHIPline (2.5°E) and in solid lines the 10-day rolling mean (i.e., low-pass filtered pCO2 where the duration is set to 10 days). We used the following single model grid cells: 42°S, 2.5°Ein the Sub-Antarctic Zone (SAZ); 44°S, 2.5°E on the Sub-Antarctic Front (SAF); and 47°S, 2.5°E in the Polar Frontal Zone (PFZ). **(b)** shows the RMSE map of the difference of the 10-day rolling mean from the daily model pCO_2 in the study domain divided by the SAF (black dashed line, Figure B.1b) into two sub-domains: the SAZ and the PFZ. This RMSE gives us a statistical understanding of what the uncertainty might be if we sampled at a 10-day rate.

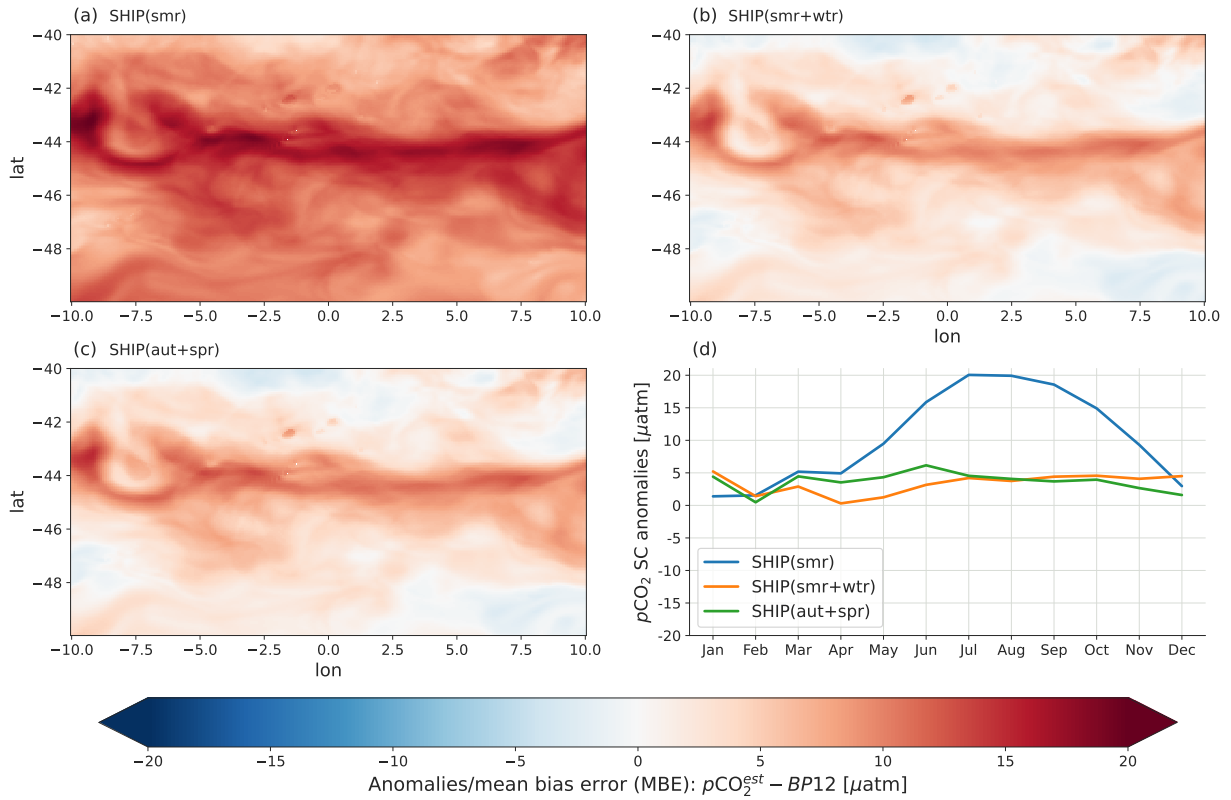


Figure B.2: Reconstruction anomalies for all the three SHIP experiments where the idealized ship sampled the domain based on the three sampling regimes, SHIP(smr) for summer, SHIP(smr+wtr) for summer and winter, and SHIP(aut+spr) for autumn and spring. **(a)**, **(b)** and **(c)** show the maps of the reconstruction anomalies from these experiements , while **(d)** shows the corresponding anomalies of the mean seasonal cycle reconstructions.

The resulting root mean square error, mean absolute error, mean bias error and Pearson correlation coefficient from the SHIP(smr+spr) experiment were 7.07 µatm, 5.5 µatm, 3.57 µatm and 0.72, respectively. Results from the SHIP(smr) and SHIP(smr+wtr) experiments were presented (Table 3.1) and discussed in the main text.

Asset C

OSSEs in the Southern Ocean as a whole

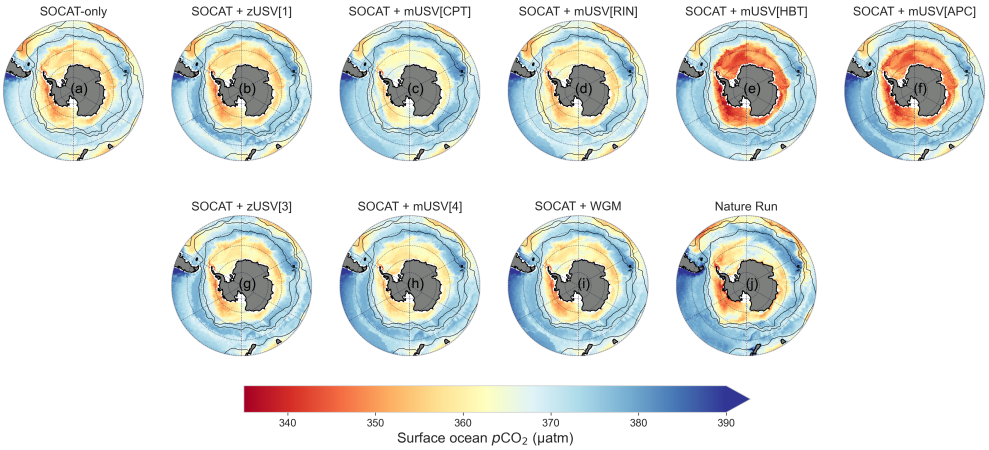


Figure C.1: Nature Run (true model) surface ocean pCO_2 **(j)** and reconstructed surface ocean pCO_2 based on the three-member ensemble average derived from SOCAT-only **(a)**, SOCAT + WGM (i) and multi-platform integrated systems **(b-h)**.

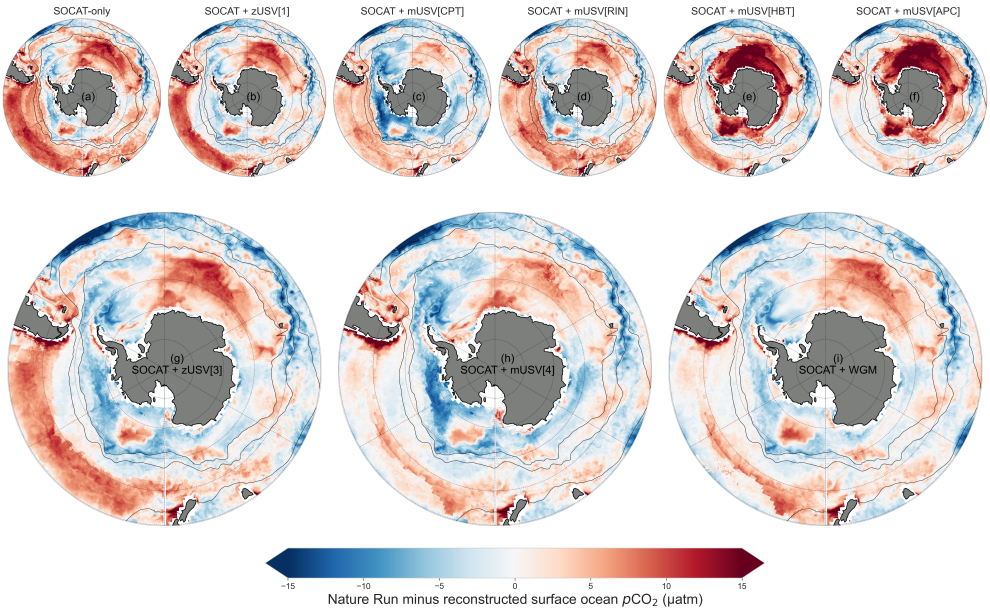


Figure C.2: Differences between the Nature Run (true model) surface ocean pCO_2 and reconstructed surface ocean pCO_2 based on the three-member ensemble average derived from SOCAT-only **(a)**, SOCAT + WGM **(i)** and multi-platform integrated systems **(b-h)**.

Bibliography

Andersson, E. and M. Masutani (2010). *Collaboration on observing system simulation experiments (Joint OSSE). ECMWF Newsletter, No. 123*. Vol. 14-16. ECMWF (cit. on p. 14).

Atlas, R. (1997). "Atmospheric observations and experiments to assess their usefulness in data assimilation". In: *J. Meteor. Soc. Japan* 75, pp. 111–130 (cit. on pp. 13, 14).

Atlas, Robert, Ross N. Hoffman, Zaizhong Ma, G. David Emmitt, Sidney A. Wood, Steven Greco, Sara Tucker, Lisa Bucci, Bachir Annane, R. Michael Hardesty, and Shirley Murillo (Sept. 2015). "Observing System Simulation Experiments (OSSEs) to Evaluate the Potential Impact of an Optical Autocovariance Wind Lidar (OAWL) on Numerical Weather Prediction". In: *Journal of Atmospheric and Oceanic Technology* 32 (9), pp. 1593–1613 (cit. on pp. 13, 14).

Aumont, O., C. Ethé, A. Tagliabue, L. Bopp, and M. Gehlen (2015). "PISCES-v2: an ocean biogeochemical model for carbon and ecosystem studies". In: *Geoscientific Model Development* 8.8, pp. 2465–2513 (cit. on p. 18).

Bakker, D. C. E., B. Pfeil, C. S. Landa, N. Metzl, K. M. O'Brien, A. Olsen, K. Smith, C. Cosca, S. Harasawa, S. D. Jones, S. Nakaoka, Y. Nojiri, U. Schuster, T. Steinhoff, C. Sweeney, T. Takahashi, B. Tilbrook, C. Wada, R. Wanninkhof, S. R. Alin, C. F. Balestrini, L. Barbero, N. R. Bates, A. A. Bianchi, F. Bonou, J. Boutin, Y. Bozec, E. F. Burger, W.-J. Cai, R. D. Castle, L. Chen, M. Chierici, K. Currie, W. Evans, C. Featherstone, R. A. Feely, A. Fransson, C. Goyet, N. Greenwood, L. Gregor, S. Hankin, N. J. Hardman-Mountford, J. Harlay, J. Hauck, M. Hoppema, M. P. Humphreys, C. W. Hunt, B. Huss, J. S. P. Ibánhez, T. Johannessen, R. Keeling, V. Kitidis, A. Körtzinger, A. Kozyr, E. Krasakopoulou, A. Kuwata, P. Landschützer, S. K. Lauvset, N. Lefèvre, C. Lo Monaco, A. Manke, J. T. Mathis, L. Merlivat, F. J. Millero, P. M. S. Monteiro, D. R. Munro, A. Murata, T. Newberger, A. M. Omar, T. Ono, K. Paterson, D. Pearce, D. Pierrot, L. L. Robbins, S. Saito, J. Salisbury, R. Schlitzer, B. Schneider, R. Schweitzer, R. Sieger, I. Skjelvan, K. F. Sullivan, S. C. Sutherland, A. J. Sutton, K. Tadokoro, M. Telszewski, M. Tuma, S. M. A. C. van Heuven, D. Vandemark, B. Ward, A. J. Watson, and S. Xu (2016). "A multi-decade record of high-quality fCO2 data in version 3 of the Surface Ocean CO2 ATlas (SOCAT)". In: *Earth System Science Data* 8.2, pp. 383–413 (cit. on pp. 4–6, 39, 40, 66, 68, 83, 91, 96, 100, 101, 105).

Béjar, Javier (2013). "K-means vs Mini Batch K-means: A comparison". In: *Departament de Llenguatges i Sistemes Informàtics, Universitat Politècnica de Catalunya* (cit. on p. 77).

Boukabara, Sid Ahmed, Isaac Moradi, Robert Atlas, Sean P.F. Casey, Lidia Cucurull, Ross N. Hoffman, Kayo Ide, V. Krishna Kumar, Ruifang Li, Zhenglong Li, Michiko Masutani, Narges Shahroudi, Jack Woollen, and Yan Zhou (Aug. 2016). "Community Global Observing System Simulation Experiment (OSSE) Package (CGOP): Description and Usage". In: *Journal of Atmospheric and Oceanic Technology* 33 (8), pp. 1759–1777 (cit. on pp. 14, 15).

Breiman, Leo (1998). “Arcing Classifiers”. In: *The Annals of Statistics* 26.3, pp. 801–824 (cit. on p. 28).

Bushinsky, Seth M., Peter Landschützer, Christian Rödenbeck, Alison R. Gray, David Baker, Matthew R. Mazloff, Laure Resplandy, Kenneth S. Johnson, and Jorge L. Sarmiento (2019). “Reassessing Southern Ocean Air-Sea CO2 Flux Estimates With the Addition of Biogeochemical Float Observations”. In: *Global Biogeochemical Cycles* 33.11, pp. 1370–1388 (cit. on pp. 5–7, 38–40, 51, 54, 56, 59, 61, 66, 90, 95, 96, 100, 105, 113, 118, 121, 122, 132).

Caldeira, Ken and Philip B Duffy (2000). “The role of the Southern Ocean in uptake and storage of anthropogenic carbon dioxide”. In: *Science* 287.5453, pp. 620–622 (cit. on p. 12).

Canadell, J.G., P.M.S. Monteiro, M.H. Costa, L. Cotrim da Cunha, P.M. Cox, A.V. Eliseev, S. Henson, M. Ishii, S. Jaccard, C. Koven, A. Lohila, P.K. Patra, S. Piao, J. Rogelj, S. Syampungani, S. Zaehle, and K. Zickfeld (2021). “Global Carbon and other Biogeochemical Cycles and Feedbacks”. In: *Climate Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change*. Ed. by V. Masson-Delmotte, P. Zhai, A. Pirani, S.L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M.I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J.B.R. Matthews, T.K. Maycock, T. Waterfield, O. Yelekçi, R. Yu, and B. Zhou. Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press, pp. 673–816 (cit. on pp. 2, 3, 5, 38).

Carter, B. R., N. L. Williams, A. R. Gray, and R. A. Feely (2016). “Locally interpolated alkalinity regression for global alkalinity estimation”. In: *Limnology and Oceanography: Methods* 14.4, pp. 268–277 (cit. on p. 7).

Carter, Lionel, I.N. McCave, and Michael J.M. Williams (2008). “Chapter 4 Circulation and Water Masses of the Southern Ocean: A Review”. In: *Antarctic Climate Evolution*. Ed. by Fabio Florindo and Martin Siegert. Vol. 8. Developments in Earth and Environmental Sciences. Elsevier, pp. 85–114 (cit. on pp. 11–13).

Chapman, Christopher C., Mary Anne Lea, Amelie Meyer, Jean Baptiste Sallée, and Mark Hindell (2020). “Defining Southern Ocean fronts and their influence on biological and physical processes in a changing climate”. In: *Nature Climate Change* 10 (3), pp. 209–219 (cit. on p. 43).

Chollet, Francois et al. (2018). *Deep learning with Python*. Vol. 361. Manning New York (cit. on p. 32).

Ciais, P., C. Sabine, G. Bala, and W. Peters (2013). “Carbon and Other Biogeochemical Cycles”. In: *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*. Ed. by T.F. Stocker, D. Qin, G.K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex, and P.M. Midgley. Cambridge University Press, pp. 465–570 (cit. on pp. 2, 3).

Denvil-Sommer, A., M. Gehlen, and M. Vrac (2021). “Observation system simulation experiments in the Atlantic Ocean for enhanced surface ocean pCO2 reconstructions”. In: *Ocean Science* 17.4, pp. 1011–1030 (cit. on pp. 15, 36, 66, 94, 105).

Denvil-Sommer, A., M. Gehlen, M. Vrac, and C. Mejia (2019). “LSCE-FFNN-v1: a two-step neural network model for the reconstruction of surface ocean pCO2 over the global ocean”. In: *Geoscientific Model Development* 12.5, pp. 2091–2105 (cit. on pp. 5, 6, 8, 9, 24, 38, 47, 93, 123, 132).

DeVries, Tim, Mark Holzer, and Francois Primeau (2017). “Recent increase in oceanic carbon uptake driven by weaker upper-ocean overturning”. In: *Nature* 542.7640, pp. 215–218 (cit. on pp. 9–13, 38, 89, 90, 97, 127).

Djeutchouang, L. M., N. Chang, L. Gregor, M. Vichi, and P. M. S. Monteiro (2022). “The sensitivity of pCO_2 reconstructions to sampling scales across a Southern Ocean sub-domain: a semi-idealized ocean sampling simulation approach”. In: *Biogeosciences* 19.17, pp. 4171–4195 (cit. on pp. 100, 112, 118).

Fay, A. R., N. S. Lovenduski, G. A. McKinley, D. R. Munro, C. Sweeney, A. R. Gray, P. Landschützer, B. B. Stephens, T. Takahashi, and N. Williams (2018). “Utilizing the Drake Passage Time-series to understand variability and change in subpolar Southern Ocean pCO_2”. In: *Biogeosciences* 15.12, pp. 3841–3855 (cit. on pp. 5, 6, 38, 39, 56).

Fay, A. R. and G. A. McKinley (2014). “Global open-ocean biomes: mean and temporal variability”. In: *Earth Syst. Sci. Data* 6, pp. 273–284 (cit. on pp. xi, xiii, 43, 49, 68, 71, 72, 74, 78, 79, 91, 92, 94, 98, 104, 123, 124).

Feizollah, Ali, Nor Badrul Anuar, Rosli Salleh, and Fairuz Amalina (2014). “Comparative Study of K-means and Mini Batch K-means Clustering Algorithms in Android Malware Detection Using Network Traffic Analysis”. In: (cit. on p. 77).

Frery, Jordan, Amaury Habrard, Marc Sebban, Olivier Caelen, and Liyun He-Guelton (2017). “Efficient Top Rank Optimization with Gradient Boosting for Supervised Anomaly Detection BT - Machine Learning and Knowledge Discovery in Databases”. In: ed. by Michelangelo Ceci, Jaakko Hollmén, Ljupčo Todorovski, Celine Vens, and Sašo Džeroski. Springer International Publishing, pp. 20–35 (cit. on p. 67).

Friedlingstein, P. et al. (2022). “Global Carbon Budget 2022”. In: *Earth System Science Data* 14.11, pp. 4811–4900 (cit. on pp. 2, 5, 11).

Friedlingstein, Pierre, Matthew W Jones, Robbie M Andrew, C E Bakker, Judith Hauck, Corinne LeQuéré, Glen P Peters, Julia Pongratz, Stephen Sitch, Josep G Canadell, Philippe Ciais, Simone R Alin, Peter Anthoni, Nicholas R Bates, Meike Becker, Laurent Bopp, Thi Tuyet Trang Chau, Frédéric Chevallier, Louise P Chini, Margot Cronin, Kim I Currie, Bertrand Decharme, Laique M Djeutchouang, Wiley Evans, Richard A Feely, Liang Feng, Thomas Gasser, Dennis Gilfillan, Francesco Tubiello, Guido van der Werf, Nicolas Vuichard, Rik Wanninkhof, Andrew J Watson, David Willis, Andrew J Wiltshire,

Wenping Yuan, Chao Yue, Xu Yue, Sönke Zaehle, and Jiye Zeng (2021). "Global Carbon Budget 2021". In: *Earth System Science Data Discussions* 2021, pp. 1–191 (cit. on pp. 2–5, 8, 9, 16, 38, 66, 90, 95, 99).

Friedman, Jerome H. (2001). "Greedy Function Approximation: A Gradient Boosting Machine". In: *The Annals of Statistics* 29.5, pp. 1189–1232 (cit. on p. 29).

Frölicher, Thomas L., Jorge L. Sarmiento, David J. Paynter, John P. Dunne, John P. Krasting, and Michael Winton (2015). "Dominance of the Southern Ocean in Anthropogenic Carbon and Heat Uptake in CMIP5 Models". In: *Journal of Climate* 28.2, pp. 862–886 (cit. on pp. 8, 9, 13, 38, 127).

Gentemann, C. L., Joel P. Scott, Piero L.F. Mazzini, Cassia Pianca, Santha Akella, Peter J. Minnett, Peter Cornillon, Baylor Fox-Kemper, Ivona Cetinić, T. Mike Chin, Jose Gomez-Valdes, Jorge Vazquez-Cuervo, Vardis Tsontos, Lisan Yu, Richard Jenkins, Sebastien De Halleux, Dave Peacock, and Nora Cohen (June 2020). "Saildrone: Adaptively Sampling the Marine Environment". In: *Bulletin of the American Meteorological Society* 101 (6), E744–E762 (cit. on pp. 45, 46, 62).

Gloege, Lucas, Galen A. McKinley, Peter Landschützer, Amanda R. Fay, Thomas L. Frölicher, John C. Fyfe, Tatiana Ilyina, Steve Jones, Nicole S. Lovenduski, Keith B. Rodgers, Sarah Schlunegger, and Yohei Takano (2021). "Quantifying Errors in Observationally Based Estimates of Ocean Carbon Sink Variability". In: *Global Biogeochemical Cycles* 35.4. e2020GB006788 2020GB006788, e2020GB006788 (cit. on pp. 5, 15, 16, 21, 37, 39, 40, 44, 47, 88, 105, 122, 129, 130, 132).

Goddijn-Murphy, L. M., D. K. Woolf, P. E. Land, J. D. Shutler, and C. Donlon (2015). "The OceanFlux Greenhouse Gases methodology for deriving a sea surface climatology of CO_2 fugacity in support of air-sea gas flux studies". In: *Ocean Science* 11.4, pp. 519–541 (cit. on p. 91).

Goodfellow, Ian, Yoshua Bengio, Aaron Courville, and Yoshua Bengio (2016). *Deep learning*. Vol. 1. 2. MIT press Cambridge (cit. on p. 30).

Grare, Laurent, Nicholas M. Statom, Nick Pizzo, and Luc Lenain (Aug. 2021). "Instrumented Wave Gliders for Air-Sea Interaction and Upper Ocean Research". In: *Frontiers in Marine Science* 8 (August), pp. 1–21 (cit. on p. 45).

Grassl, Hartmut (2001). "Chapter 1.1 Climate and oceans". In: *Ocean Circulation and Climate*. Ed. by Gerold Siedler, John Church, and John Gould. Vol. 77. International Geophysics. Academic Press, pp. 3–9 (cit. on p. 11).

Gray, Alison R., Kenneth S. Johnson, Seth M. Bushinsky, Stephen C. Riser, Joellen L. Russell, Lynne D. Talley, Rik Wanninkhof, Nancy L. Williams, and Jorge L. Sarmiento (2018). "Autonomous Biogeochemical Floats Detect Significant Carbon Dioxide Outgassing in the High-Latitude Southern Ocean". In: *Geophysical Research Letters* 45.17, pp. 9049–9057 (cit. on pp. 6, 8, 39, 40, 43, 45, 56, 59, 61, 66, 69, 90, 94, 100, 102, 121).

Gregor, L. and N. Gruber (2021). "OceanSODA-ETHZ: a global gridded data set of the surface ocean carbonate system for seasonal to decadal studies of ocean acidification". In: *Earth System Science Data* 13.2, pp. 777–808 (cit. on pp. 37, 39, 43, 47, 57, 66, 91, 93, 123, 132).

Gregor, Luke (Mar. 2017). "Improved estimates and understanding of interannual trends of CO2 fluxes in the Southern Ocean". Doctoral dissertation. University of Cape Town (cit. on p. 19).

Gregor, Luke, Schalk Kok, and Pedro M. S. Monteiro (2018). "Interannual drivers of the seasonal cycle of CO2 in the Southern Ocean". In: *Biogeosciences* 15.8, pp. 2361–2378 (cit. on pp. 9, 10, 15, 21, 34, 38, 43, 54, 55, 57, 66, 88, 95, 100, 105, 115, 116).

— (2017). "Empirical methods for the estimation of Southern Ocean CO2: Support vector and random forest regression". In: *Biogeosciences* 14 (23), pp. 5551–5569 (cit. on pp. 5, 15, 16, 20, 21, 34, 38, 81, 95, 104, 112).

Gregor, Luke, Alice D. Lebehot, Schalk Kok, and Pedro M. Scheel Monteiro (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?" In: *Geosci. Model Dev.* 12.12, pp. 5113–5136 (cit. on pp. xiii, 5–9, 16, 20, 21, 24, 34, 38, 40, 41, 43, 44, 47, 48, 51, 53, 54, 56, 57, 66, 72, 74, 86, 90–94, 97, 100, 104, 112, 117, 118, 121–123, 126, 129, 132).

Gruber, Nicolas, Dorothee CE Bakker, Tim DeVries, Luke Gregor, Judith Hauck, Peter Landschützer, Galen A McKinley, and Jens Daniel Müller (2023). "Trends and variability in the ocean carbon sink". In: *Nature Reviews Earth & Environment*, pp. 1–16 (cit. on p. 3).

Gruber, Nicolas, Manuel Gloor, Sara E. Mikaloff Fletcher, Scott C. Doney, Stephanie Dutkiewicz, Michael J. Follows, Markus Gerber, Andrew R. Jacobson, Fortunat Joos, Keith Lindsay, Dimitris Menemenlis, Anne Mouchet, Simon A. Müller, Jorge L. Sarmiento, and Taro Takahashi (2009). "Oceanic sources, sinks, and transport of atmospheric CO2". In: *Global Biogeochemical Cycles* 23.1, pp. 1–21 (cit. on pp. 13, 22).

Gruber, Nicolas, Peter Landschützer, and Nicole S. Lovenduski (2019). "The Variable Southern Ocean Carbon Sink". In: *Annual Review of Marine Science* 11.1, pp. 159–186 (cit. on pp. 3, 9–13, 22, 38, 40, 90, 115, 121).

Halliwell, G. R., A. Srinivasan, V. Kourafalou, H. Yang, D. Willey, M. Le Hénaff, and R. Atlas (2014). "Rigorous Evaluation of a Fraternal Twin Ocean OSSE System for the Open Gulf of Mexico". In: *Journal of Atmospheric and Oceanic Technology* 31.1, pp. 105–130 (cit. on pp. 13, 14).

Hartigan, J. A. and M. A. Wong (1979). "Algorithm AS 136: A K-Means Clustering Algorithm". In: *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 28.1, pp. 100–108 (cit. on p. 75).

Hastie, Trevor, Robert Tibshirani, and Jerome Friedman (2009). *The elements of statistical learning: data mining, inference, and prediction*. Springer Science and Business Media (cit. on pp. 24, 31, 32).

Hauck, J., C. Völker, D. A. Wolf-Gladrow, C. Laufkötter, M. Vogt, O. Aumont, L. Bopp, E. T. Buitenhuis, S. C. Doney, J. Dunne, N. Gruber, T. Hashioka, J. John, C. LeQuéré, I. D. Lima, H. Nakano, R. Séférian, and I. Totterdell (2015). "On the Southern Ocean CO2 uptake and the role of the biological carbon pump in the 21st century". In: *Global Biogeochemical Cycles* 29 (9), pp. 1451–1470 (cit. on p. 57).

Hauck, Judith, Moritz Zeising, Corinne LeQuéré, Nicolas Gruber, Dorothee C.E. Bakker, Laurent Bopp, Thi Tuyet Trang Chau, Özgür Gürses, Tatiana Ilyina, Peter Landschützer, Andrew Lenton, Laure Resplandy, Christian Rödenbeck, Jörg Schwinger, and Roland Séférian (Oct. 2020). "Consistency and Challenges in the Ocean Carbon Sink Estimate for the Global Carbon Budget". In: *Frontiers in Marine Science* 7, p. 852 (cit. on pp. 3, 8, 16, 38, 57, 66, 90, 94, 95, 99, 104, 130).

Hicks, Stephanie C., Ruoxi Liu, Yuwei Ni, Elizabeth Purdom, and Davide Risso (Jan. 2021). "mbkmeans: Fast clustering for single cell data using mini-batch k-means". In: *PLOS Computational Biology* 17.1, pp. 1–18 (cit. on p. 77).

Hine, R, S Willcox, G Hine, and T Richardson (2009). "The Wave Glider: A Wave-Powered autonomous marine vehicle". In: IEEE, pp. 1–6 (cit. on p. 45).

Hoffman, Ross N. and Robert Atlas (Sept. 2016). "Future Observing System Simulation Experiments". In: *Bulletin of the American Meteorological Society* 97 (9), pp. 1601–1616 (cit. on p. 14).

Holte, James, Lynne D. Talley, John Gilson, and Dean Roemmich (2017). "An Argo mixed layer climatology and database". In: *Geophysical Research Letters* 44.11, pp. 5618–5626 (cit. on p. 21).

Iida, Yosuke, Atsushi Kojima, Yusuke Takatani, Toshiya Nakano, Hiroyuki Sugimoto, Takashi Midorikawa, and Masao Ishii (Dec. 2015). "Trends in pCO2 and sea-air CO2 flux over the global open oceans for the last two decades". In: *Journal of Oceanography* 71 (6), pp. 637–661 (cit. on p. 15).

Ishii, Masao, Hisayuki Y. Inoue, Hidekazu Matsueda, and Eiichiro Tanoue (1998). "Close coupling between seasonal biological production and dynamics of dissolved inorganic carbon in the Indian Ocean sector and the western Pacific Ocean sector of the Antarctic Ocean". In: *Deep Sea Research Part I: Oceanographic Research Papers* 45.7, pp. 1187–1209 (cit. on p. 115).

Jacobson, Andrew R., Sara E. Mikaloff-Fletcher, Nicolas Gruber, Jorge L. Sarmiento, and Manuel Gloor (2007). "A joint atmosphere-ocean inversion for surface fluxes of carbon dioxide: 1. Methods and global-scale fluxes". In: *Global Biogeochemical Cycles* 21.1 (cit. on p. 3).

Jain, Anil K. (2010). "Data clustering: 50 years beyond K-means". In: *Pattern Recognition Letters* 31.8. Award winning papers from the 19th International Conference on Pattern Recognition (ICPR), pp. 651–666 (cit. on p. 75).

Johnson, Kenneth S., Joshua N. Plant, Luke J. Coletti, Hans W. Jannasch, Carole M. Sakamoto, Stephen C. Riser, Dana D. Swift, Nancy L. Williams, Emmanuel Boss, Nils Haëntjens, Lynne D. Talley, and Jorge L. Sarmiento (Aug. 2017). "Biogeochemical sensor performance in the SOCCOM profiling float array". In: *Journal of Geophysical Research: Oceans* 122 (8), pp. 6416–6436 (cit. on p. 45).

Jones, S. D., C. Le Quéré, and C. Rödenbeck (2012). "Autocorrelation characteristics of surface ocean pCO2 and air-sea CO2 fluxes". In: *Global Biogeochemical Cycles* 26.2 (cit. on p. 127).

Jones, Steve D., Corinne LeQuéré, Christian Rödenbeck, Andrew C. Manning, and Are Olsen (2015). "A statistical gap-filling method to interpolate global monthly surface ocean carbon dioxide data". In: *Journal of Advances in Modeling Earth Systems* 7.4, pp. 1554–1575 (cit. on p. 15).

Kamenkovich, Igor, Angelique Haza, Alison R. Gray, Carolina O. Dufour, and Zulema Garraffo (2017). "Observing System Simulation Experiments for an array of autonomous biogeochemical profiling floats in the Southern Ocean". In: *Journal of Geophysical Research: Oceans* 122.9, pp. 7595–7611 (cit. on pp. 13, 14, 36).

Ke, Guolin, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, and Tie-Yan Liu (2017). "Lightgbm: A highly efficient gradient boosting decision tree". In: *Advances in neural information processing systems*, pp. 3146–3154 (cit. on pp. 28, 29, 47, 132).

Keppler, Lydia and Peter Landschützer (2019). "Regional Wind Variability Modulates the Southern Ocean Carbon Sink". In: *Scientific reports* 9.1, p. 7384 (cit. on pp. 8–10, 43, 93).

Khatiwala, S., T. Tanhua, S. Mikaloff-Fletcher, M. Gerber, S. C. Doney, H. D. Graven, N. Gruber, G. A. McKinley, A. Murata, A. F. Ríos, and C. L. Sabine (2013). "Global ocean storage of anthropogenic carbon". In: *Biogeosciences* 10.4, pp. 2169–2191 (cit. on p. 3).

Landschützer, P, N Gruber, DCE Bakker, and U Schuster (2014). "Recent variability of the global ocean carbon sink". In: *Global Biogeochemical Cycles* 28.9, pp. 927–949 (cit. on pp. 15, 16, 38, 39, 43, 66, 72, 93, 123, 129).

Landschützer, P, N Gruber, DCE Bakker, U Schuster, S Nakaoka, MR Payne, TP Sasse, and J Zeng (2013). "A neural network-based estimate of the seasonal to inter-annual variability of the Atlantic Ocean carbon sink". In: *Biogeosciences* 10.11, pp. 7793–7815 (cit. on pp. 5, 15, 20, 24).

Landschützer, Peter, Nicolas Gruber, and Dorothee C. E. Bakker (2016). "Decadal variations and trends of the global ocean carbon sink". In: *Global Biogeochem. Cycles* 30.10, pp. 1396–1417 (cit. on pp. 5, 6, 8, 16, 38, 40, 43, 44, 47, 55, 66, 104, 123, 126, 129, 132).

Landschützer, Peter, Nicolas Gruber, F Alexander Haumann, Christian Rödenbeck, Dorothee CE Bakker, Steven Van Heuven, Mario Hoppema, Nicolas Metzl, Colm Sweeney, Taro Takahashi, et al. (2015). "The reinvigoration of the Southern Ocean carbon sink". In: *Science* 349.6253, pp. 1221–1224 (cit. on pp. 8–11, 15, 38, 90, 115).

Lenton, Andrew, Richard J. Matear, and Bronte Tilbrook (Dec. 2006). "Design of an observational strategy for quantifying the Southern Ocean uptake of CO2". In: *Global Biogeochemical Cycles* 20 (4), p. 4010 (cit. on pp. 4, 5, 54, 56, 58, 63, 118, 129).

Lenton, Andrew, Bronte Tilbrook, R Law, Dorothee CE Bakker, Scott C Doney, Nicolas Gruber, Mario Hoppema, M Ishii, Nicole S Lovenduski, Richard J Matear, et al. (2013). "Sea-air CO2 fluxes in the

Southern Ocean for the period 1990-2009". In: *Biogeosciences Discussions* 10, pp. 285–333 (cit. on pp. 4, 9, 11, 13, 15, 54, 91, 103).

LeQuéré, Corinne, Robbie M Andrew, Josep G Canadell, Stephen Sitch, Jan Ivar Korsbakken, Glen P Peters, Andrew C Manning, Thomas A Boden, Pieter P Tans, Richard A Houghton, Ralph F Keeling, Simone Alin, Oliver D Andrews, Peter Anthoni, Leticia Barbero, Laurent Bopp, Frédéric Chevallier, Louise P. Chini, Philippe Ciais, Kim Currie, Christine Delire, Scott C. Doney, Pierre Friedlingstein, Thanos Gkritzalis, Ian Harris, Judith Hauck, Vanessa Haverd, Mario Hoppema, Kees Klein Goldewijk, Atul K. Jain, Etsushi Kato, Arne Körtzinger, Peter Landschützer, Nathalie Lefèvre, Andrew Lenton, Sebastian Lienert, Danica Lombardozzi, Joe R. Melton, Nicolas Metzl, Frank Millero, Pedro M. S. Monteiro, David R. Munro, Julia E. M. S. Nabel, Shin-ichiro Nakaoka, Kevin O'Brien, Are Olsen, Abdirahman M. Omar, Tsuneo Ono, Denis Pierrot, Benjamin Poulter, Christian Rödenbeck, Joe Salisbury, Ute Schuster, Jörg Schwinger, Roland Séférian, Ingunn Skjelvan, Benjamin D. Stocker, Adrienne J. Sutton, Taro Takahashi, Hanqin Tian, Bronte Tilbrook, Ingrid T. van der Laan-Luijkx, Guido R. van der Werf, Nicolas Viovy, Anthony P. Walker, Andrew J. Wiltshire, and Sönke Zaehle (2016). "Global Carbon Budget 2016". In: *Earth Syst. Sci. Data* 8.2, pp. 605–649 (cit. on p. 16).

LeQuéré, Corinne, Christian Rödenbeck, Erik T Buitenhuis, Thomas J Conway, Ray Langenfelds, Antony Gomez, Casper Labuschagne, Michel Ramonet, Takakiyo Nakazawa, Nicolas Metzl, Nathan Gillett, and Martin Heimann (2007). "Sink Due to Recent Climate Change". In: *Science* 316.June, pp. 1735–1738 (cit. on pp. 9, 11, 38).

Li, Lijuan, Yongqiang Yu, Yanli Tang, Pengfei Lin, Jinbo Xie, Mirong Song, Li Dong, Tianjun Zhou, Li Liu, Lu Wang, Ye Pu, Xiaolong Chen, Lin Chen, Zhenghui Xie, Hongbo Liu, Lixia Zhang, Xin Huang, Tao Feng, Weipeng Zheng, Kun Xia, Hailong Liu, Jiping Liu, Yan Wang, Longhuan Wang, Binghao Jia, Feng Xie, Bin Wang, Shuwen Zhao, Zipeng Yu, Bowen Zhao, and Jilin Wei (2020). "The Flexible Global Ocean-Atmosphere-Land System Model Grid-Point Version 3 (FGOALS-g3): Description and Evaluation". In: *Journal of Advances in Modeling Earth Systems* 12.9. e2019MS002012 2019MS002012, e2019MS002012 (cit. on p. 16).

Lovenduski, Nicole S., Nicolas Gruber, and Scott C. Doney (2008). "Toward a mechanistic understanding of the decadal trends in the Southern Ocean carbon sink". In: *Global Biogeochemical Cycles* 22.3, pp. 1–9 (cit. on pp. 35, 74, 78, 83, 110).

Majkut, Joseph D., Brendan R. Carter, Thomas L. Frölicher, Carolina O. Dufour, Keith B. Rodgers, and Jorge L. Sarmiento (2014). "An observing system simulation for Southern Ocean carbon dioxide uptake". In: *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 372.2019, p. 20130046 (cit. on pp. 15, 36, 38–40, 44, 113).

Maritorena, Stéphane, Odile Hembise Fanton d Andon, Antoine Mangin, and David A. Siegel (2010). "Merged satellite ocean color data products using a bio-optical model: Characteristics, benefits and issues". In: *Remote Sensing of Environment* 114.8, pp. 1791–1804 (cit. on p. 21).

Marshall, Gareth J (2003). "Trends in the Southern Annular Mode from observations and reanalyses". In: *Journal of climate* 16.24, pp. 4134–4143 (cit. on p. 11).

Marshall, John and Kevin Speer (2012). "Closure of the meridional overturning circulation through Southern Ocean upwelling". In: *Nature geoscience* 5.3, pp. 171–180 (cit. on pp. 11, 13).

Mazloff, M. R., B. D. Cornuelle, S. T. Gille, and A. Verdy (Feb. 2018). "Correlation Lengths for Estimating the Large-Scale Carbon and Heat Content of the Southern Ocean". In: *Journal of Geophysical Research: Oceans* 123 (2), pp. 883–901 (cit. on pp. 14, 36, 113).

McCarty, Will, David Carvalho, Isaac Moradi, and Nikki C. Privé (2021). "Observing System Simulation Experiments Investigating Atmospheric Motion Vectors and Radiances from a Constellation of 4 – 5-μm Infrared Sounders". In: *Journal of Atmospheric and Oceanic Technology* 38.2, pp. 331–347 (cit. on p. 36).

McKinley, Galen A., Amanda R. Fay, Yassir A. Eddebbar, Lucas Gloege, and Nicole S. Lovenduski (June 2020). "External Forcing Explains Recent Decadal Variability of the Ocean Carbon Sink". In: *AGU Advances* 1 (2), e2019AV000149 (cit. on pp. 9, 10, 38).

McNeil, Ben I. and Richard J. Matear (2008). "Southern Ocean acidification: A tipping point at 450-ppm atmospheric CO2". In: *Proceedings of the National Academy of Sciences* 105.48, pp. 18860–18864 (cit. on p. 9).

Meinig, Christian, Eugene F. Burger, Nora Cohen, Edward D. Cokelet, Meghan F. Cronin, Jessica N. Cross, Sebastien De Halleux, Richard Jenkins, Andrew T. Jessup, Calvin W. Mordy, Noah Lawrence-Slavas, Adrienne J. Sutton, Dongxiao Zhang, and Chidong Zhang (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". In: *Frontiers in Marine Science* 6 (JUL), pp. 1–15 (cit. on pp. 45, 46, 102–104, 113, 118).

Meinig, Christian, Noah Lawrence-Slavas, Richard Jenkins, and Heather M. Tabisola (Feb. 2015). "The use of Saildrones to examine spring conditions in the Bering Sea: Vehicle specification and mission performance". In: Institute of Electrical and Electronics Engineers Inc., pp. 1–6 (cit. on p. 62).

— (Feb. 2016). "The use of Saildrones to examine spring conditions in the Bering Sea: Vehicle specification and mission performance". In: *OCEANS 2015 - MTS/IEEE Washington* (cit. on p. 45).

Metzl, N., C. Brunet, A. Jabaud-Jan, A. Poisson, and B. Schauer (2006). "Summer and winter air-sea CO2 fluxes in the Southern Ocean". In: *Deep Sea Research Part I: Oceanographic Research Papers* 53.9, pp. 1548–1563 (cit. on p. 13).

Mongwe, N. Precious, Nicolette Chang, and Pedro M.S. Monteiro (2016). "The seasonal cycle as a mode to diagnose biases in modelled CO2 fluxes in the Southern Ocean". In: *Ocean Modelling* 106, pp. 90–103 (cit. on pp. 19, 22, 44, 48).

Mongwe, N. Precious, Marcello Vichi, and Pedro M. S. Monteiro (2018). "The seasonal cycle of pCO2 and CO2 fluxes in the Southern Ocean: diagnosing anomalies in CMIP5 Earth system models". In: *Biogeosciences* 15.9, pp. 2851–2872 (cit. on pp. 16, 22, 41, 44, 54, 88, 90–92, 95).

Monteiro, Pedro M.S., Luke Gregor, Marina Lévy, Stacy Maenner, Christopher L. Sabine, and Sebastiaan Swart (Oct. 2015). "Intraseasonal variability linked to sampling alias in air-sea CO2 fluxes in the Southern Ocean". In: *Geophysical Research Letters* 42 (20), pp. 8507–8514 (cit. on pp. 22, 37, 39, 40, 43, 54, 56, 59, 61, 91, 100, 101, 103, 115, 129).

Monteiro, Pedro M.S., Ute Schuster, Maria Hood, Andrew Lenton, Nicolas Metzl, Are Olsen, Keith Rogers, Chris Sabine, Taro Takahashi, Bronte Tilbrook, James Yoder, Rik Wanninkhof, and Andrew J. Watson (Dec. 2010). "A Global Sea Surface Carbon Observing System: Assessment of Changing Sea Surface CO2 and Air-Sea CO2 Fluxes". In: European Space Agency, pp. 702–714 (cit. on pp. 5–7, 22, 56, 58, 63, 69, 103).

Moore, J Keith, Mark R Abbott, James G Richman, Walker O Smith, Timothy J Cowles, Kenneth H Coale, Wilford D Gardner, and Richard T Barber (1999). "SeaWiFS satellite ocean color data from the Southern Ocean". In: *Geophysical Research Letters* 26.10, pp. 1465–1468 (cit. on pp. 35, 74).

Munro, David R., Nicole S. Lovenduski, Taro Takahashi, Britton B. Stephens, Timothy Newberger, and Colm Sweeney (2015). "Recent evidence for a strengthening CO2 sink in the Southern Ocean from carbonate system measurements in the Drake Passage (2002-2015)". In: *Geophysical Research Letters* 42.18, pp. 7623–7630 (cit. on pp. 48, 56).

Natekin, Alexey and Alois Knoll (2013). "Gradient boosting machines, a tutorial". In: *Frontiers in Neurorobotics* 7 (cit. on pp. 24, 28, 132).

Nevison, C. D., M. Manizza, R. F. Keeling, B. B. Stephens, J. D. Bent, J. Dunne, T. Ilyina, M. Long, L. Resplandy, J. Tjiputra, and S. Yukimoto (2016). "Evaluating CMIP5 ocean biogeochemistry and Southern Ocean carbon uptake using atmospheric potential oxygen: Present-day performance and future projection". In: *Geophysical Research Letters* 43.5, pp. 2077–2085 (cit. on p. 11).

Nicholson, Sarah-Anne, Marina Lévy, Joan Llort, Sebastiaan Swart, and Pedro MS Monteiro (2016). "Investigation into the impact of storms on sustaining summer primary productivity in the Sub-Antarctic Ocean". In: *Geophysical Research Letters* 43.17, pp. 9192–9199 (cit. on p. 91).

Nicholson, Sarah-Anne, Daniel B. Whitt, Ilker Fer, Marcel D. du Plessis, Alice D. Lebéhot, Sebastiaan Swart, Adrienne J. Sutton, and Pedro M. S. Monteiro (Jan. 2022). "Storms drive outgassing of CO2 in the subpolar Southern Ocean". In: *Nature Communications 2022 13:1* 13 (1), pp. 1–12 (cit. on pp. 6, 7, 39, 101, 129).

Orsi, Alejandro H., Thomas Whitworth, and Worth D. Nowlin (May 1995). “On the meridional extent and fronts of the Antarctic Circumpolar Current”. In: *Deep Sea Research Part I: Oceanographic Research Papers* 42 (5), pp. 641–673 (cit. on pp. 9, 42, 103, 104).

Pfeil, B., A. Olsen, D. C.E. Bakker, S. Hankin, H. Koyuk, A. Kozyr, J. Malczyk, A. Manke, N. Metzl, C. L. Sabine, J. Akl, S. R. Alin, N. Bates, R. G.J. Bellerby, A. Borges, J. Boutin, P. J. Brown, W. J. Cai, F. P. Chavez, A. Chen, C. Cosca, A. J. Fassbender, R. A. Feely, M. González-Dávila, C. Goyet, B. Hales, N. Hardman-Mountford, C. Heinze, M. Hood, M. Hoppema, C. W. Hunt, D. Hydes, M. Ishii, T. Johannessen, S. D. Jones, R. M. Key, A. Körtzinger, P. Landschützer, S. K. Lauvset, N. Lefévre, A. Lenton, A. Lourantou, L. Merlivat, T. Midorikawa, L. Mintrop, C. Miyazaki, A. Murata, A. Nakadate, Y. Nakano, S. Nakaoka, Y. Nojiri, A. M. Omar, X. A. Padin, G. H. Park, K. Paterson, F. F. Perez, D. Pierrot, A. Poisson, A. F. Ríos, J. M. Santana-Casiano, J. Salisbury, V. V.S.S. Sarma, R. Schlitzer, B. Schneider, U. Schuster, R. Sieger, I. Skjelvan, T. Steinhoff, T. Suzuki, T. Takahashi, K. Tedesco, M. Telszewski, H. Thomas, B. Tilbrook, J. Tjiputra, D. Vandemark, T. Veness, R. Wanninkhof, A. J. Watson, R. Weiss, C. S. Wong, and H. Yoshikawa-Inoue (Apr. 2013). “A uniform, quality controlled Surface Ocean CO2 Atlas (SOCAT)”. In: *Earth System Science Data* 5 (1), pp. 125–143 (cit. on p. 39).

Prentice, IC, GD Farquhar, MJR Fasham, ML Goulden, M Heimann, VJ Jaramillo, HS Kheshgi, C LeQuéré, RJ Scholes, and DWR Wallace (2001). “The carbon cycle and atmospheric carbon dioxide”. In: *Climate change 2001: The Scientific Basis. Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change (IPCC)*. Ed. by JT Houghton, Y Ding, DJ Griggs, M Noguer, PJ van der Linden, X Dai, K Maskell, and CA Johnson. United Kingdom: Cambridge University Press (cit. on p. 4).

Resplandy, L., R. F. Keeling, C. Rödenbeck, B. B. Stephens, S. Khatiwala, K. B. Rodgers, M. C. Long, L. Bopp, and P. P. Tans (July 2018). “Revision of global carbon fluxes based on a reassessment of oceanic and riverine carbon transport”. In: *Nature Geoscience* 11 (7), pp. 504–509 (cit. on pp. 3, 115).

Rintoul, Stephen R (2008). “The role of Southern Ocean in past, present and future climate: A strategy for the International Polar Year”. In: (cit. on p. 12).

Rintoul, Stephen R., Chris W. Hughes, and Dirk Olbers (2001). “Chapter 4.6 The antarctic circumpolar current system”. In: *Ocean Circulation and Climate*. Ed. by Gerold Siedler, John Church, and John Gould. Vol. 77. International Geophysics. Academic Press, pp. 271–XXXVI (cit. on pp. 9, 11, 12).

Ritter, R., P. Landschützer, N. Gruber, A. R. Fay, Y. Iida, S. Jones, S. Nakaoka, G.-H. Park, P. Peylin, C. Rödenbeck, K. B. Rodgers, J. D. Shutler, and J. Zeng (2017). “Observation-Based Trends of the Southern Ocean Carbon Sink”. In: *Geophysical Research Letters* 44.24, pp. 12, 339–12, 348 (cit. on p. 39).

Rödenbeck, C., D. C. E. Bakker, N. Metzl, A. Olsen, C. Sabine, N. Cassar, F. Reum, R. F. Keeling, and M. Heimann (2014). “Interannual sea-air CO_2 flux variability from an observation-driven ocean mixed-layer scheme”. In: *Biogeosciences* 11.17, pp. 4599–4613 (cit. on pp. 5, 15, 35).

Rödenbeck, C., S. Houweling, M. Gloor, and M. Heimann (2003). "CO_2 flux history 1982-2001 inferred from atmospheric data using a global inversion of atmospheric transport". In: *Atmospheric Chemistry and Physics* 3.6, pp. 1919–1964 (cit. on p. 15).

Rödenbeck, Christian, Dorothee CE Bakker, Nicolas Gruber, Yosuke Iida, Andrew R Jacobson, S Jones, Peter Landschützer, Nicolas Metzl, Shin-ichiro Nakaoka, Are Olsen, et al. (2015). "Data-based estimates of the ocean carbon sink variability–first results of the Surface Ocean pCO2 Mapping intercomparison (SOCOM)". In: *Biogeosciences* 12, pp. 7251–7278 (cit. on pp. 5, 6, 8, 20, 21, 34, 39, 40, 66, 72, 93, 95, 132).

Rousseeuw, Peter J. (1987). "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis". In: *Journal of Computational and Applied Mathematics* 20, pp. 53–65 (cit. on pp. 80, 92).

Sabine, C. L., S. Hankin, H. Koyuk, D. C. E. Bakker, B. Pfeil, A. Olsen, N. Metzl, A. Kozyr, A. Fassbender, A. Manke, J. Malczyk, J. Akl, S. R. Alin, R. G. J. Bellerby, A. Borges, J. Boutin, P. J. Brown, W.-J. Cai, F. P. Chavez, A. Chen, C. Cosca, R. A. Feely, M. González-Dávila, C. Goyet, N. Hardman-Mountford, C. Heinze, M. Hoppema, C. W. Hunt, D. Hydes, M. Ishii, T. Johannessen, R. M. Key, A. Körtzinger, P. Landschützer, S. K. Lauvset, N. Lefèvre, A. Lenton, A. Lourantou, L. Merlivat, T. Midorikawa, L. Mintrop, C. Miyazaki, A. Murata, A. Nakadate, Y. Nakano, S. Nakaoka, Y. Nojiri, A. M. Omar, X. A. Padin, G.-H. Park, K. Paterson, F. F. Perez, D. Pierrot, A. Poisson, A. F. RıI AM HERE!!!!os, J. Salisbury, J. M. Santana-Casiano, V. V. S. S. Sarma, R. Schlitzer, B. Schneider, U. Schuster, R. Sieger, I. Skjelvan, T. Steinhoff, T. Suzuki, T. Takahashi, K. Tedesco, M. Telszewski, H. Thomas, B. Tilbrook, D. Vandemark, T. Veness, A. J. Watson, R. Weiss, C. S. Wong, and H. Yoshikawa-Inoue (2013). "Surface Ocean CO_2 Atlas (SOCAT) gridded data products". In: *Earth System Science Data* 5.1, pp. 145–153 (cit. on pp. 4, 39, 40).

Sabine, Christopher, Adrienne Sutton, Kelly McCabe, Noah Lawrence-Slavas, Simone Alin, Richard Feely, Richard Jenkins, Stacy Maenner, Christian Meinig, Jesse Thomas, Erik van Ooijen, Abe Passmore, and Bronte Tilbrook (Aug. 2020). "Evaluation of a New Carbon Dioxide System for Autonomous Surface Vehicles". In: *Journal of Atmospheric and Oceanic Technology* 37 (8), pp. 1305–1317 (cit. on p. 45).

Sabine, Christopher L., Richard A. Feely, Nicolas Gruber, Robert M. Key, Kitack Lee, John L. Bullister, Rik Wanninkhof, C. S. Wong, Douglas W. R. Wallace, Bronte Tilbrook, Frank J. Millero, Tsung-Hung Peng, Alexander Kozyr, Tsueno Ono, and Aida F. Rios (2004). "The Oceanic Sink for Anthropogenic CO2". In: *Science* 305.5682, pp. 367–371 (cit. on pp. 2, 9, 12, 38).

Sallée, J.-B., E. Shuckburgh, N. Bruneau, A. J. S. Meijers, T. J. Bracegirdle, Z. Wang, and T. Roy (2013). "Assessment of Southern Ocean water mass circulation and characteristics in CMIP5 models: Historical bias and forcing response". In: *Journal of Geophysical Research: Oceans* 118.4, pp. 1830–1844 (cit. on pp. 13, 127).

Sarmiento, J. L., R. Slater, R. Barber, L. Bopp, S. C. Doney, A. C. Hirst, J. Kleypas, R. Matear, U. Mikolajewicz, P. Monfray, V. Soldatov, S. A. Spall, and R. Stouffer (2004). "Response of ocean ecosystems to climate warming". In: *Global Biogeochemical Cycles* 18.3 (cit. on p. 13).

Sarmiento, Jorge L and Nicolas Gruber (2006). *Ocean biogeochemical dynamics*. Princeton and Oxford: Princeton University Press (cit. on pp. 9–11, 24).

Sigman, Daniel M., Mathis P. Hain, and Gerald H. Haug (July 2010). "The polar ocean and glacial cycles in atmospheric CO2 concentration". In: *Nature 2010 466:7302* 466 (7302), pp. 47–55 (cit. on p. 11).

Sutton, A. J., N. L. Williams, and B. Tilbrook (2021). "Constraining Southern Ocean CO2 Flux Uncertainty Using Uncrewed Surface Vehicle Observations". In: *Geophysical Research Letters* 48.3, e2020GL091748 (cit. on pp. 6, 39, 40, 45–47, 51, 54, 56, 59, 61, 62, 100–104, 113, 118).

Swart, Sebastiaan, Nicolette Chang, Nicolas Fauchereau, Warren Joubert, Mike Lucas, Thato Mtshali, Alakendra Roychoudhury, Alessandro Tagliabue, Sandy Thomalla, Howard Waldron, and Pedro M.S. Monteiro (2012). "Southern Ocean Seasonal Cycle Experiment 2012: Seasonal scale climate and carbon cycle links". In: *South African Journal of Science* 108 (3-4), pp. 3–5 (cit. on p. 22).

Sweeney, Colm, Emanuel Gloor, Andrew R Jacobson, Robert M Key, Galen McKinley, Jorge L Sarmiento, and Rik Wanninkhof (2007). "Constraining global air-sea gas exchange for CO2 with recent bomb 14C measurements". In: *Global Biogeochemical Cycles* 21.2 (cit. on p. 4).

Takahashi, Taro, Jon Olafsson, John G. Goddard, David W. Chipman, and S. C. Sutherland (1993). "Seasonal variation of CO2 and nutrients in the high-latitude surface oceans: A comparative study". In: *Global Biogeochemical Cycles* 7.4, pp. 843–878 (cit. on pp. 16, 34, 38, 41, 75).

Takahashi, Taro, Stewart C. Sutherland, Rik Wanninkhof, Colm Sweeney, Richard A. Feely, David W. Chipman, Burke Hales, Gernot Friederich, Francisco Chavez, Christopher Sabine, Andrew Watson, Dorothee C.E. Bakker, Ute Schuster, Nicolas Metzl, Hisayuki Yoshikawa-Inoue, Masao Ishii, Takashi Midorikawa, Yukihiro Nojiri, Arne Körtzinger, Tobias Steinhoff, Mario Hoppema, Jon Olafsson, Thorarinn S. Arnarson, Bronte Tilbrook, Truls Johannessen, Are Olsen, Richard Bellerby, C.S. Wong, Bruno Delille, N.R. Bates, and Hein J.W. de Baar (2009). "Climatological mean and decadal change in surface ocean pCO2, and net sea-air CO2 flux over the global oceans". In: *Deep Sea Research Part II: Topical Studies in Oceanography* 56.8. Surface Ocean CO2 Variability and Vulnerabilities, pp. 554–577 (cit. on pp. 9, 11, 43, 54, 90, 91, 95, 97).

Talley, L. D., I. Rosso, I. Kamenkovich, M. R. Mazloff, J. Wang, E. Boss, A. R. Gray, K. S. Johnson, R. M. Key, S. C. Riser, N. L. Williams, and J. L. Sarmiento (2019). "Southern Ocean Biogeochemical Float Deployment Strategy, With Example From the Greenwich Meridian Line (GO-SHIP A12)". In: *Journal of Geophysical Research: Oceans* 124 (1), pp. 403–431 (cit. on pp. 46, 104).

Talley, Lynne D. (Mar. 2013). “Closure of the Global Overturning Circulation Through the Indian, Pacific, and Southern Oceans: Schematics and Transports”. In: *Oceanography* 26.1, pp. 80–97 (cit. on pp. 11, 12).

Taylor, John R., Scott Bachman, Megan Stamper, Phil Hosegood, Katherine Adams, Jean-Baptiste Sallee, and Ricardo Torres (2018). “Submesoscale Rossby waves on the Antarctic circumpolar current”. In: *Science Advances* 4.3, eaao2824 (cit. on p. 127).

Terhaar, Jens, Thomas L. Frölicher, and Fortunat Joos (2021). “Southern Ocean anthropogenic carbon sink constrained by sea surface salinity”. In: *Science Advances* 7.18, eabd5964 (cit. on pp. 11, 12).

Thomalla, S. J., N. Fauchereau, S. Swart, and P. M.S. Monteiro (2011). “Regional scale characteristics of the seasonal cycle of chlorophyll in the Southern Ocean”. In: *Biogeosciences* 8 (10), pp. 2849–2866 (cit. on pp. 63, 72, 75, 91, 92, 95, 115).

Valsala, Vinu, M.G. Sreeush, M. Anju, Pentakota Sreenivas, Yogesh K. Tiwari, Kunal Chakraborty, and S. Sijikumar (2021). “An observing system simulation experiment for Indian Ocean surface pCO2 measurements”. In: *Progress in Oceanography* 194, p. 102570 (cit. on pp. 15, 36).

Vichi, Marcello, J. Icarus Allen, Simona Masina, and Nicholas J. Hardman-Mountford (June 2011). “The emergence of ocean biogeochemical provinces: A quantitative assessment and a diagnostic for model evaluation”. In: *Global Biogeochemical Cycles* 25 (2), p. 2005 (cit. on p. 72).

Wanninkhof, Rik (1992). “Relationship between wind speed and gas exchange over the ocean”. In: *Journal of Geophysical Research: Oceans* 97.C5, pp. 7373–7382 (cit. on p. 4).

— (2014). “Relationship between wind speed and gas exchange over the ocean revisited”. In: *Limnology and Oceanography: Methods* 12.6, pp. 351–362 (cit. on p. 4).

Watson, Andrew J., Ute Schuster, Jamie D. Shutler, Thomas Holding, Ian G.C. Ashton, Peter Landschützer, David K. Woolf, and Lonneke Goddijn-Murphy (Dec. 2020). “Revised estimates of ocean-atmosphere CO2 flux are consistent with ocean carbon inventory”. In: *Nature Communications* 11 (1) (cit. on pp. 3, 5, 66).

Whitt, Christopher, Jay Pearlman, Brian Polagye, Frank Caimi, Frank Muller-Karger, Andrea Copping, Heather Spence, Shyam Madhusudhana, William Kirkwood, Ludovic Grosjean, Bilal Muhammad Fiaz, Satinder Singh, Sikandra Singh, Dana Manalang, Ananya Sen Gupta, Alain Maguer, Justin J. H. Buck, Andreas Marouchos, Malayath Aravindakshan Atmanand, Ramasamy Venkatesan, Vedachalam Narayanaswamy, Pierre Testor, Elizabeth Douglas, Sebastien de Halleux, and Siri Jodha Khalsa (2020). “Future Vision for Autonomous Ocean Observations”. In: *Frontiers in Marine Science* 7 (cit. on p. 102).

Williams, N. L., L. W. Juranek, R. A. Feely, K. S. Johnson, J. L. Sarmiento, L. D. Talley, A. G. Dickson, A. R. Gray, R. Wanninkhof, J. L. Russell, S. C. Riser, and Y. Takeshita (2017). “Calculating surface ocean pCO2 from biogeochemical Argo floats equipped with pH: An uncertainty analysis”. In: *Global Biogeochemical Cycles* 31.3, pp. 591–604 (cit. on pp. 6–8, 39, 40, 44, 56, 59, 99).

Wu, Yingxu, Mathis P. Hain, Matthew P. Humphreys, Sue Hartman, and Toby Tyrrell (2019). “What drives the latitudinal gradient in open-ocean surface dissolved inorganic carbon concentration?” In: *Biogeosciences* 16 (13), pp. 2661–2681 (cit. on pp. 57, 89).

Xue, Ming, Mingjing Tong, and Kelvin K. Droegemeier (2006). “An OSSE Framework Based on the Ensemble Square Root Kalman Filter for Evaluating the Impact of Data from Radar Networks on Thunderstorm Analysis and Forecasting”. In: *Journal of Atmospheric and Oceanic Technology* 23.1, pp. 46–66 (cit. on p. 13).

Zeng, J., T. Matsunaga, N. Saigusa, T. Shirai, S. Nakaoka, and Z. -H. Tan (2017). “Technical note: Evaluation of three machine learning models for surface ocean CO2 mapping”. In: *Ocean Science* 13.2, pp. 303–313 (cit. on pp. 34, 132).

Zeng, Jiye, Yukihiro Nojiri, Shin-ichiro Nakaoka, Hideaki Nakajima, and Tomoko Shirai (2015). “Surface ocean CO2 in 1990-2011 modelled using a feed-forward neural network”. In: *Geoscience Data Journal* 2.1, pp. 47–51 (cit. on pp. 5, 34).

Zeng, Xubin, Robert Atlas, Ronald J. Birk, Frederick H. Carr, Matthew J. Carrier, Lidia Cucurull, William H. Hooke, Eugenia Kalnay, Raghu Murtugudde, Derek J. Posselt, Joellen L. Russell, Daniel P. Tyndall, Robert A. Weller, and Fuqing Zhang (2020). “Use of Observing System Simulation Experiments in the United States”. In: *Bulletin of the American Meteorological Society* 101.8, E1427–E1438 (cit. on pp. 13, 14, 36, 37).