An Evaluation of Global Ocean Models in the North Atlantic using BGC-Argo Observations

by

Melina Mehlmann

Submitted in partial fulfilment of the requirements

for the degree of Master of Science

at

Dalhousie University

Halifax, Nova Scotia

March 2024

Dalhousie University is located in Mi'kma'ki, the ancestral and unceded territory of the Mi'kmaq. We are all Treaty people.

TABLE OF CONTENTS

LIST OF TABLES ..... **v**

LIST OF FIGURES ..... **vi**

ABSTRACT ..... **ix**

LIST OF ABBREVIATIONS USED ..... **x**

1 INTRODUCTION **1**

1.1 North Atlantic Circulation and Biogeochemistry *1*

1.2 Previous Ocean Model Evaluations *3*

1.3 (BGC-)Argo Floats *5*

1.4 World Ocean Atlas (WOA) *8*

1.5 Objectives *8*

1.6 Analytical Framework *9*

2 DATA AND METHODS **11**

2.1 Models *11*

2.2 Observations *13*

2.3 Variables *14*

2.4 BGC-Provinces *15*

2.4 Interpolation Methods *16*

2.5 Standard Deviation (STD) *17*

2.6 Misfit *18*

2.7 Bias splitting *18*

3 RESULTS **20**

3.1 Introduction *20*

3.2 How do float observations compare to WOA observations? *20*

3.3 Do global ocean models compare well with observations? *22*

3.3.1 Surface Comparison 22

3.3.2 Annual Cycles in the northern North Atlantic 27

3.3.3 Physical properties in the North Atlantic 28

3.4 Are global ocean models in good agreement within the natural variability of the observations? *29*

3.5 Are BGC biases linked to biases in physical properties? *30*

3.6 Do individual global ocean models outperform a CMIP6 ensemble? *33*

4 DISCUSSION 35

4.1. Evaluating whether global ocean models can reproduce an accurate state of North Atlantic properties. *35*

4.1.1 Underestimated phytoplankton blooms in the northern North Atlantic 36

4.1.2 Biases in ocean physics 37

4.1.3 Misplacement of the Gulf Stream 38

4.1.4 Does a good agreement of CHL in the mixed layer imply a similar agreement throughout the entire euphotic zone? 39

4.1.5 Is euphotic zone CHL affected by the representation of NO_3 below the euphotic zone? 39

4.1.6 Is the representation of ocean properties within the euphotic zone the same as it is below? 40

4.2 Investigate the interrelation between biases in BGC and physical models. *40*

4.2.1 Does a misfit in the model physics affect agreement of BGC variables? 40

4.2.2 Does a more complex BGC model or increased spatial resolution guarantee better model performance? 41

4.3 Assessment of a CMIP6 model ensemble performance with respect to individual models including data-assimilation models. *43*

4.4. Uncertainties *44*

5 CONCLUSION **46**

APPENDICES **49**

A.1 Longhurst province processing *49*

A.2 Δ-Profile Comparisons *52*

A.3 Annually Averaged Surface Biases and Surface Concentrations *55*

A.4 Bias Correlation *58*

A.5 AOU Mismatch Test *60*

REFERENCES **61**

LIST OF TABLES

Table 1: Overview of models included in evaluation. 12

Table 2: Overview of BGC complexity of BGC models included in evaluation. 12

Table 3: Model-Argo bias overview. 47

Table 4: Number of profiles used in this evaluation per property and province. Corresponding profile locations in Figure 15. 51

LIST OF FIGURES

Figure 1: North Atlantic circulation including Gulf Stream (GS), Azores Current (AC), North Atlantic Current (NAC), Iceland Current (IC), Labrador Current (LC), and Deep Western Boundary Current (DWBC). Different coloured arrows explained in legend. Regions of deep and intermediate water formation indicated by purple and blue circles. Figure adapted from Puerta et al. (2020).....2

Figure 2: Schematic description of the cycle of an Argo float. *Design by Thomas Haessig* (https://argo.ucsd.edu/how-do-floats-work/). 7

Figure 3: World map displaying 3908 (March 2024) active Argo float locations (https://argo.ucsd.edu/about/status/). 7

Figure 4: Adjusted BGC-provinces with "good Argo coverage" in the North Atlantic. ARCT: Arctic, SARC: Southern Arctic, NADR: North Atlantic Drift, GFST: Gulf Stream, NASW: North Atlantic Subtropical Gyre, NATR: North Atlantic Tropical Gyre, WTRA: Western tropical Atlantic.....16

Figure 5: 0-110m averaged annual cycles of northern North Atlantic provinces ARCT (top panels), SARC (middle panels), and NADR (bottom panels) for properties CHL, NO_3, AOU, SAL & TEMP. Argo data in dashed black, WOA in dashed blue, MOI in light blue dotted line, and CMIP6 models in remaining solid colours. WOA, MOI, and CMIP6 data from locations of Argo profiles only. Grey shading displays ±1 standard deviation of Argo data. 21

Figure 6: Average delta (model-Argo) depth profiles for CHL, NO_3, OXY, SAL & TEMP for provinces NADR (northern North Atlantic) and WTRA (southern North Atlantic). Black x with ±1 standard deviation indicates satellite CHL, Argo data in dashed black, WOA in dashed blue, MOI in light blue dotted line, CMIP6 models in remaining solid colours, and the black dotted line is an ensemble of all CMIP6 models (ALL). WOA, MOI, and CMIP6 data from locations of Argo profiles only. Grey shading displays ±1 standard deviation of Argo data. CHL profiles shown to 250m depth, other properties to 2000m. 22

Figure 7: Annual surface NO_3 [µmol/kg] average. Top row left panel shows WOA annual average from 1900-2017. Top row second from the left is data-assimilating model MOI from 2009-2021. Remaining panels are CMIP6 models, averaged over 1990-2019. Surface refers to an average over 0-6m. 23

Figure 8: Annual average surface (0-6m) TEMP bias (Model-WOA) maps. Top row left panel shows WOA annual average from 1900-2017. Top row second from the left is data-assimilating model MOI from 2009-2021. Remaining panels are CMIP6 models, averaged over 1990-2019. 24

Figure 9: Annual average (500-1200) TEMP bias (Model-WOA) maps. Top row left panel shows WOA annual average from 1900-2017. Top row second from the left is data-

assimilating model MOI from 2009-2021. Remaining panels are CMIP6 models, averaged over 1990-2019. 25

Figure 10: TEMP biases vs. NO_3 biases at 500-1200m for provinces NADR, GFST, NASW & NATR. Coloured lines indicate regression for each model, Pearson's correlation coefficients r are given in legend. Negative r indicates anticorrelation. .31

Figure 11: NO_3 bias analysis, splitting the NO_3 bias (left panels) into a component based on ocean physics (preformed NO_3 bias: NO_3^0, vertical middle panels) and a bias component due to BGC processes (NO_3^{bio}, right panels) for three example provinces ARCT (top panels), NATR (horizontal middle panels), and WTRA (bottom panels). Solid coloured lines refer to CMIP6 model biases, dotted blue line indicates MOI data-assimilation model. 32

Figure 12: Top: Model misfits for each province (rows) for each property (panels). ALL denotes CMIP6 ensemble mean. Bottom: Average model rank based on lowest misfit for Euphotic Zone (EZ) (0-100 m), below Euphotic Zone (BEZ) (130-2000 m), and full profile (0-2000 m). Lowest misfit/rank indicates best model. ALL is excluded given that individual models outperform ensemble mean. 34

Figure 13: Original Longhurst provinces in the North Atlantic. 49

Figure 14: Left panel shows provinces that were adjusted/combined due to (BGC-)Argo coverage and shelves. Right panel shows adjusted provinces used in the evaluation. 49

Figure 15: Argo profile locations. Every panel displays profile locations for a different property. See Table 4 for number of profiles. 50

Figure 16: Histogram of number of month with more than five good quality-flagged (1/2/5/8) Argo profiles per property (different colours, see legend) for each province. Provinces that exceed the threshold are WTRA, NATR, NASW, GFST, NADR, SARC & ARCT. Provinces marked in red are left out of the evaluation. 51

Figure 17: For all 5 figures above. Average delta (model-Argo) depth profiles for CHL, NO_3, OXY, SAL & TEMP, each figure for a different province. Black x with ±1 standard deviation indicates satellite CHL, Argo data in dashed black, WOA in dashed blue, MOI in light blue dotted line, CMIP6 models in remaining solid colours, and the black dotted line is an ensemble of all CMIP6 models (ALL). WOA, MOI, and CMIP6 data from locations of Argo profiles only. Grey shading displays ±1 standard deviation of Argo data. CHL profiles shown to 250m depth, other properties to 2000m. 54

Figure 18: Annually averaged surface OXY [µmol/kg] bias (Model-WOA). Top row left panel shows WOA annual average from 1900-2017. Top row second from the left is

data-assimilating model MOI from 2009-2021. Remaining panels are CMIP6 models, averaged over 1990-2019. Surface refers to an average over 0-6m. 55
Figure 19: Annually averaged surface SAL bias (Model-WOA). Top row left panel shows WOA annual average from 1900-2017. Top row second from the left is data-assimilating model MOI from 2009-2021. Remaining panels are CMIP6 models, averaged over 1990-2019. Surface refers to an average over 0-6m. 56

Figure 20: Annual surface CHL [mg/m^3] average. Top row left panel shows satellite annual average from 2003-2021. Top row second from the left is data-assimilating model MOI from 2009-2021. Remaining panels are CMIP6 models, averaged over 1990-2019. Surface model data refers to an average over 0-6m. 57

Figure 21: TEMP biases vs. NO_3 biases at 500-1200m for provinces ARCT, SARC, NADR & GFST. Coloured lines indicate regression for each model, Pearson's correlation coefficients r are given in legend. Negative r indicates anticorrelation. 58

Figure 22: OXY biases vs. NO_3 biases at 500-1200m for provinces NADR, GFST, NASW & NATR. Coloured lines indicate regression for each model, Pearson's correlation coefficients r are given in legend. Negative r indicates anticorrelation. 59

Figure 23: Test explaining mismatch between MOI and (BGC-)Argo AOU in provinces ARCT, SARC & NADR (panels from left to right). AOU was calculated from MOI OXY, TEMP & SAL and then compared to calculations using (BGC-)Argo data for either OXY, TEMP or SAL or both TEMP and OXY. Percentages listed indicate how much of the mismatch is being explained by each property (e.g. MOI AOU calculations using Argo OXY explain 65-71% of the mismatch between Argo and MOI AOU). Percentages were calculated form averages over 0-110m depth and were further averaged over time which leads to some rounding errors. 60

ABSTRACT

Testing, evaluating, and enhancing global ocean models holds paramount significance in climate and oceanographic research, particularly for projecting climate change impacts. While the evaluation of physical model components has become increasingly comprehensive and sophisticated, an evaluation of the biogeochemical (BGC) components has so far been difficult to accomplish because of a lack of extensive global-scale BGC observations. In this thesis, North Atlantic (BGC-)Argo profile data of chlorophyll-a, nitrate, and oxygen from the surface to 2000 m depth are compared to the corresponding properties simulated by state-of-the-art global ocean models of the CMIP6 ensemble, as well as Mercator Ocean's data assimilating models MOI-BIO4 and MOI-GLO12. The physical variables salinity and temperature are included to investigate relationships between model physics and BGC. World Ocean Atlas (WOA) climatologies serve as a secondary dataset for model assessment, as well as to validate the methodology of utilizing (BGC-)Argo data for model evaluation and further to conduct comparisons with prior studies that have emphasized on WOA analyses.

To enable a comprehensive assessment of model-data misfits and the description of model biases, the North Atlantic is divided into BGC-provinces defined by unique environmental features. Results indicate large misfits between CMIP6 model properties and observations, specifically nitrate underestimation of up to 10 µmol/kg within the euphotic zone and up to 20 µmol/kg at intermediate depths (500-1200m). Modelled temperatures discrepancies are found for surface and intermediate waters in the Labrador Sea, Greenland Sea, Norwegian Sea, and the Gulf Stream separation region, for which biases of up to 5°C are reported. Data assimilating models are in better agreement with the observations than CMIP6 models.

This study suggests that neither heightened BGC complexity nor increased model resolution independently guarantee improved performance.

LIST OF ABBREVIATIONS USED

Biogeochemical Provinces:

ARCT: Atlantic Arctic

BPLR: Boreal Polar

CARB: Caribbean

CNRY: Canary current coast

ETRA: Eastern tropical Atlantic

GFST: Gulf Stream

GUIA: Guianas coast

GUIN: Guinea current coast

MEDI: Mediterranean Sea

NADR: North Atlantic Drift

NASE: Northeast Atlantic subtropical gyral

NASW: Northwest Atlantic subtropical gyral

NATR: North Atlantic tropical gyral

NECS: Northeast Atlantic shelves

NWCS: Northwest Atlantic shelves

SARC: Atlantic sub-Arctic

WTRA: Western tropical Atlantic

Properties:

AOU: Seawater Apparent Oxygen Utilization

CHL: Seawater chlorophyll-a concentration

NO_3: Seawater nitrate concentration

OXY: Seawater oxygen concentration

SAL: Seawater salinity

TEMP: Seawater temperature

Models:

CanESM5: Canadian Centre for Climate Modelling and Analysis Earth System Model 5

CMCC-ESM4: Centro Euro-Mediterraneo sui Cambiamenti Climatici Earth System Model 4

GFDL-ESM4: Geophysical Fluid Dynamics Laboratory Earth System Model 4
IPSL-CM6A-LR: Institute Pierre-Simon Laplace Climate Model 6A Low Resolution
MPI-ESM1.2-HR: Max Planck Institute Earth System Model 1.2 High Resolution
MOI-BIO4: Mercator Ocean International Biogeochemical Model ¼° Resolution
MOI-GLO12: Mercator Ocean International Glorys $^{1}/_{12}$° Resolution
NorESM5: Norwegian Meteorological Institute Earth System Model 5

Other Abbreviations and Symbols:

AR: Assessment Report
BCP: Biological carbon pump
BGC: Biogeochemical
CDR: Carbon Dioxide Removal
CMIP6: Coupled Model Intercomparison Project 6
Δ: Difference Model minus Observation
ESM: Earth System Model
IPCC: Intergovernmental Panel on Climate Change
NAC: North Atlantic Current
SSP: Shared socioeconomic pathways
WCRP: World Climate Research Programme
WOA: World Ocean Atlas

ACKNOWLEDGEMENTS

I would like to begin by acknowledging that Dalhousie University is located in Mi'kma'ki, the ancestral and unceded territory of the Mi'kmaq. We are all Treaty people.

First and foremost, I extend my deepest gratitude to my supervisor Dr. Katja Fennel for her unwavering support, honesty, and encouragement throughout this journey. Your guidance and insights have been invaluable in navigating the contours of my research and finding the path that best aligns with my priorities.

I want to thank the MEMG lab for listening to many terribly prepared presentations and for your kind feedback that always resulted in an improvement of my work.

I am also immensely grateful to my committee for their support and flexibility as I pursued an unconventional route to graduation. Their understanding and encouragement have been instrumental in reaching this milestone.

To my friends, especially Lina, Jess, Jenna, May, Ruby, Marie, Britton, Benji, Emma, Emmanuelle & Lydia, your continuous support and willingness to listen to me vent has been a constant source of strength and motivation. I am truly blessed to have you in my life.

To my family, I owe a debt of gratitude that words cannot express. To my mother-in-law Leslie, thank you for being the best grandmother our daughter could have wished for and for picking up our daughter from daycare every single day. And to my husband, Philip, who tirelessly cared for our daughter in the evenings, allowing me to focus on my work, even if it meant working late into the night for himself. Your sacrifices and support mean everything to me.

I would like to express my appreciation to the Dalhousie Oceanography Graduate Student Association (DOSA) for their invaluable student support. This community has made my graduate studies a wonderful memory that I will cherish forever.

Furthermore, I am deeply thankful to Cusanuswerk for their generous financial support, without which pursuing this master's degree would not have been possible.

I also want to extend my gratitude to all individuals involved in the creation and improvement of the models I utilized, as well as those involved in the data I analyzed. Your contributions have been integral to the success of my research.

Lastly, I want to acknowledge the importance of data access, whether it be observations or models, in advancing scientific understanding. Your contributions to data accessibility are crucial for future research endeavors.

1 INTRODUCTION

1.1 North Atlantic Circulation and Biogeochemistry

The North Atlantic's circulation patterns, characterized by a double gyre system influenced by large-scale winds (Sverdrup, 1947, Stommel, 1948, Munk, 1950), exhibit distinctive features such as upwelling and deep-water formation sites (Pickart et al., 2002; Dickson et al., 2002). The currents in the North Atlantic constitute a complex and influential system of oceanic circulation with the North Atlantic Current (NAC) as a key component. Originating as a continuation of the Gulf Stream, the NAC extends across the Atlantic Ocean, following a northward trajectory along the eastern seaboard of North America before turning eastward towards Europe. As part of the North Atlantic subtropical gyre, the NAC plays a critical role in redistributing heat from lower to higher latitudes. Its warm and relatively saline waters impact regional climate patterns, notably contributing to the temperate conditions experienced by Western Europe. The NAC interacts dynamically with other currents, such as the Iceland Current and the Labrador Current (Figure 1), creating a complex system of eddies and meanders that impact oceanic ecosystems by influencing nutrient availability and determining the distribution patterns of marine species (Krauss et al., 1987; Rossby, 1996, Carr & Rossby, 2001).

The Gulf Stream, a strong, warm ocean current in the Atlantic Ocean, separates from the continental shelf of North America. As the Gulf Stream flows northward along the eastern coast of the United States, it follows the continental shelf until it reaches approximately 35°N, typically around Cape Hatteras, North Carolina. At this point, known as the Gulf Stream separation point, the Gulf Stream detaches from the continental shelf and enters the deeper waters of the North Atlantic Ocean. This separation often results in the formation of meanders and eddies as the current continues its journey across the Atlantic (Schoonover et al., 2017).

The ocean is known to be the largest active carbon dioxide sink, absorbing around 30% of anthropogenic carbon dioxide (Gruber et al., 2019, Sabine et al., 2004) through marine carbon pumps including the solubility and biological carbon pump (BCP) (Volk & Hoffert, 1985). The BCP is a major component of the global carbon cycle: Organic carbon is being

transferred from the sunlit surface waters to the deep ocean, mostly as sinking particles (Sigman and Boyle, 2000, Boyd and Trull, 2007). Estimates state that without the BCP, atmospheric carbon dioxide concentrations would be approximately 50% higher (Parekh et al., 2006). Because of the intricate circulation patterns in the North Atlantic that play a crucial role in redistributing nutrients and the subduction of organic material, the North Atlantic BCP has a significant influence on global ocean biogeochemistry (Sanders et al., 2014). However, it is important to note, that while the BCP can influence atmospheric CO_2 concentrations over long timescales, it is not the primary mechanisms responsible for absorbing anthropogenic CO_2 emissions. Instead, physical and chemical processes (e.g. air-sea gas exchange and CO_2 dissolution) of the solubility pump play a significant role in short-time anthropogenic atmospheric CO_2 uptake by the ocean (Volk & Hoffert, 1985).

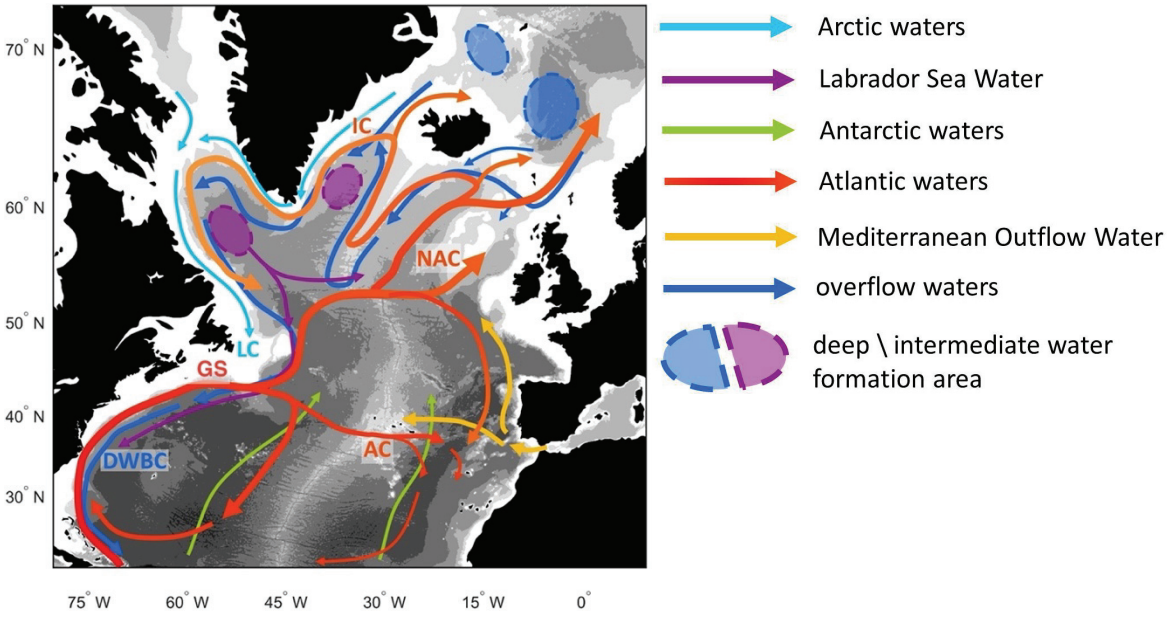


Figure 1: North Atlantic circulation including Gulf Stream (GS), Azores Current (AC), North Atlantic Current (NAC), Iceland Current (IC), Labrador Current (LC), and Deep Western Boundary Current (DWBC). Different coloured arrows explained in legend. Regions of deep and intermediate water formation indicated by purple and blue circles. Figure adapted from Puerta et al. (2020).

The springtime phytoplankton bloom in temperate and subpolar latitudes across the North Atlantic Ocean ranks among the most productive phenomena in the World Ocean. Like other productive regions, the North Atlantic's spring bloom marks a key biogeochemical

event, notably contributing to the downward transport of organic particles into the deep ocean which is fundamental to long-term carbon sequestration (Pommier et al., 2009; Siegel et al., 2012, Honjo et al., 1996). In contrast to the highly productive northern region, the southern part of the North Atlantic is characterized by the dominance of the central oligotrophic gyre, enveloping the subtropics and exhibiting a predominantly aseasonal nature (Laws et al., 2000; Vantrepotte and Mélin, 2009). Productive coastal zones south of 45°N are concentrated along the eastern margins, attributable to the upward transport of nutrients facilitated by upwelling (Brandt et al., 2023).

1.2 Previous Ocean Model Evaluations

Global ocean models are essential components of the climate models used by the International Panel of Climate Change (IPCC) to project future climate. Evaluation of model deviations from reality and description of model biases are needed to quantify model uncertainty, guide model improvement, and enable critical interpretation of model results. This is especially important when models are used with influential purposes like policy making (e.g., CMIP models, Masson-Delmotte et al., 2021 (IPCC6)).

CMIP6, a key international multi-model research initiative (Meehl et al., 2000, 2007; Taylor et al., 2012) organized by the World Climate Research Programme (WCRP), underpins climate change assessments and reports. CMIP6 reported findings and peer-reviewed publications provide the basis of assessments and reports notably contributing to the IPCC's sixth Assessment Report (AR6) (Eyring et al., 2016). CMIP6 carbon emission scenarios are based on a matrix that uses shared socioeconomic pathways (SSPs) which cover different future global societal developments ranging from no consideration of climate change over adaptation strategies to mitigation of climate change (O'Neill et al., 2015; Riahi et al., 2017).

Previous studies have focussed on the comparison of CMIP6 performance with respect to its predecessor CMIP5. As such, Séférian et al. (2020) have found that for observational assessments, most CMIP6 models generally surpass their CMIP5 counterparts in numerous regions and across various marine biogeochemical fields. Fu et al. (2022) agree on

improvements in the multi-model mean from CMIP5 to CMIP6 for 10 out of 13 properties, including oxygen and nitrate, and Richter et al. (2020) report that the CMIP5-persistent equatorial Atlantic warm SST bias has been eliminated in most CMIP6 models.

The growing availability of ocean biogeochemical data, which has contributed to an enhanced comprehension of the underlying processes, facilitates advancements in the marine biogeochemical components of the present generation of Earth System Models (ESMs) in CMIP6. This has led to marine BGC models employed in CMIP6 exhibiting greater diversity compared to those in CMIP5 (Séférian et al., 2020). Nevertheless, Séférian et al. (2020) also found systematic model-data errors, such as oxygen concentrations at 150 meters (signature depth for oxygen minimum zones) in the tropical Atlantic, that persist even in CMIP6 models. The authors state that this error might be a systematic bias in ocean BGC models which is independent from ocean resolution or complexity of the BGC model.

An evaluation study of CMIP5 and CMIP6 models by Laurent et al. (2021) reports persisting biases in CMIP6 such as underestimations of surface nitrate and chlorophyll, and further reveal differences in surface temperature and salinity indicative of spatial mismatches in the models' large-scale current systems.

Global climate models with a horizontal resolution of 1° often display errors in the representation of western boundary currents by overshooting the separation latitude (Kiehl & Gent, 2004). This leads to sea surface temperature (SST) biases which further affect a correct representation of air-sea heat fluxes (Chassignet & Marshall, 2008).

Particularly difficult to model is the Gulf Stream separation, which is a known challenge for numerical models due to its sensitivity to sub-grid scale parameterization (Bryan et al., 2007; Schoonover et al., 2017; Schoonover et al., 2016).

Increasing the horizontal resolution to 10km (1/10° at the equator) or less has been reported to show promising improvements in the numerical representation of western boundary currents separation (Chassignet & Marshall, 2008). It has been argued that at this scale, separation can occur correctly because the critical Reynolds number (defines the transition from laminar to turbulent flow) is exceeded (Dengg, 1993). Baroclinic instability processes are also well captured due to the first baroclinic Rossby radius of deformation (determines

the scale of motion for geostrophic equilibrium: Coriolis force and the pressure gradient force are in balance) being resolved for a large part of the domain (Bryan et al., 2007).

Coarse spatial resolution (0.5-2° at the equator) also prohibits an accurate portrayal of the BGC interactions between shelf and open ocean (Bonan and Doney, 2018; Holt et al., 2017) which affects the certainty of future projections (Laurent et al., 2021). Ocean margins and shelves in particular, are important components of the global ocean carbon cycle, as they host over 40% of carbon burial (Ducklow and McCallister, 2004) and account for over 30% of primary production (Muller- Karger, 2005).

However, analyses of model mean state performance in relation to model properties, including resolution and complexity, have indicated that neither increasing resolution nor complexity guarantees automatic model improvement. Instead, model enhancement is a nuanced result of improved ocean physical processes and a more accurate representation of BGC processes (Séférian et al., 2020, Laurent et al., 2021).

A different path to improved model performance is through data-assimilation. Data-assimilation aims to attain the most accurate representation of past, present, or future oceanic condition through a technique involving the statistical integration of models and observational data (Fennel et al., 2022). While many agree that data-assimilation is the most efficient approach with regards to reliable ocean state representation (Hoteit et al., 2018; Ghil and Malanotte-Rizzoli, 1991; Wunsch, 1996; Bennett, 2005), disadvantages of data-assimilation persist including increased associated computational cost and complexity.

1.3 (BGC-)Argo Floats

Argo floats are autonomous instruments deployed worldwide to collect data on ocean temperature and salinity profiles, some with additional sensors for dissolved oxygen and other properties. Since the launch of the international Argo program in 1999, these floats are an essential component of modern oceanographic research, facilitating the systematic monitoring of the world's oceans with unprecedented detail and coverage. Operating

autonomously, Argo floats descend to depths of around 2,000 meters, gathering measurements as they ascend to the surface (Figure 2). Upon reaching the surface, they transmit the collected data via satellite in real-time, providing researchers with invaluable insights into oceanographic processes and climate variability. With thousands of floats deployed across the globe, the Argo program enables continuous monitoring of ocean conditions, contributing crucial data for climate research, ocean modeling, and marine resource management (Figure 3). Argo floats represent a significant advancement in our ability to understand and monitor the dynamic and complex ocean environment, playing a vital role in addressing pressing environmental challenges and informing policies for sustainable ocean management (Jayne et al., 2017; Roemmich et al., 2009; Roemmich et al., 2019).

The BGC-Argo fleet offers a new global data set of 6 essential ocean BGC properties from the surface to 2000 m depth (Stoer et al., 2023). BGC-Argo observations have previously been used for management of ocean and global resources, as well as for improving our understanding of marine BGC processes, with which new perspectives can be employed to mitigate model limitations (Claustre et al., 2020).

In this thesis, (BGC-)Argo float data are the main observational set used for global ocean model evaluations. Modelled properties are compared to float observations at the location of Argo profiles and for the same depth levels (Argo depth are interpolated, see 2.4 Interpolation Methods).

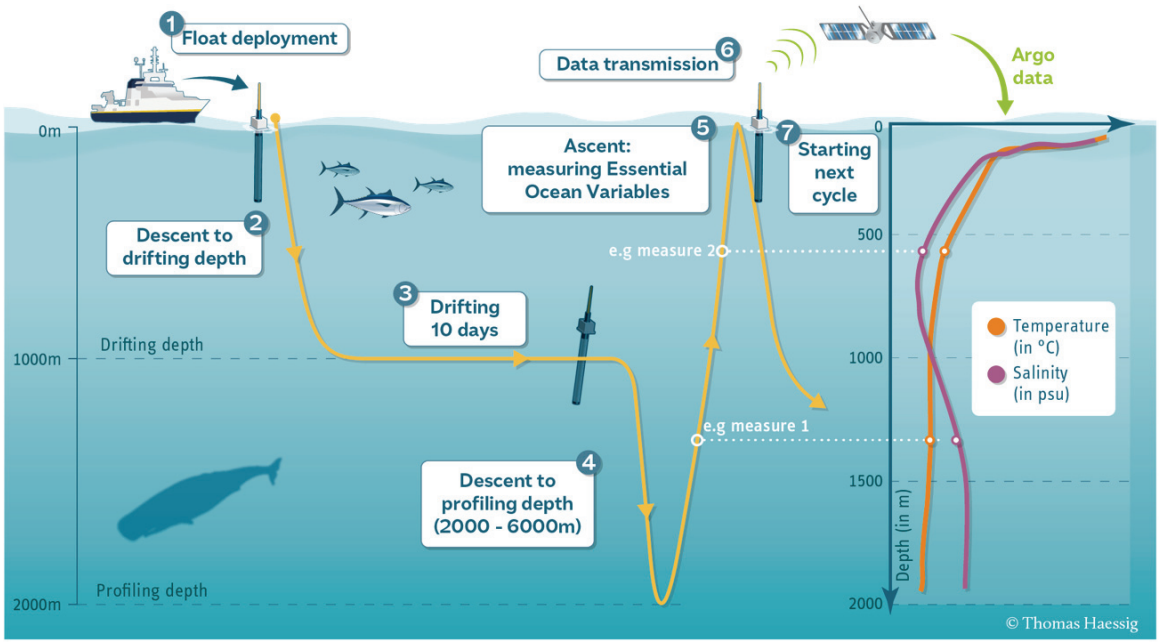


Figure 2: Schematic description of the cycle of an Argo float. *Design by Thomas Haessig* (https://argo.ucsd.edu/how-do-floats-work/).

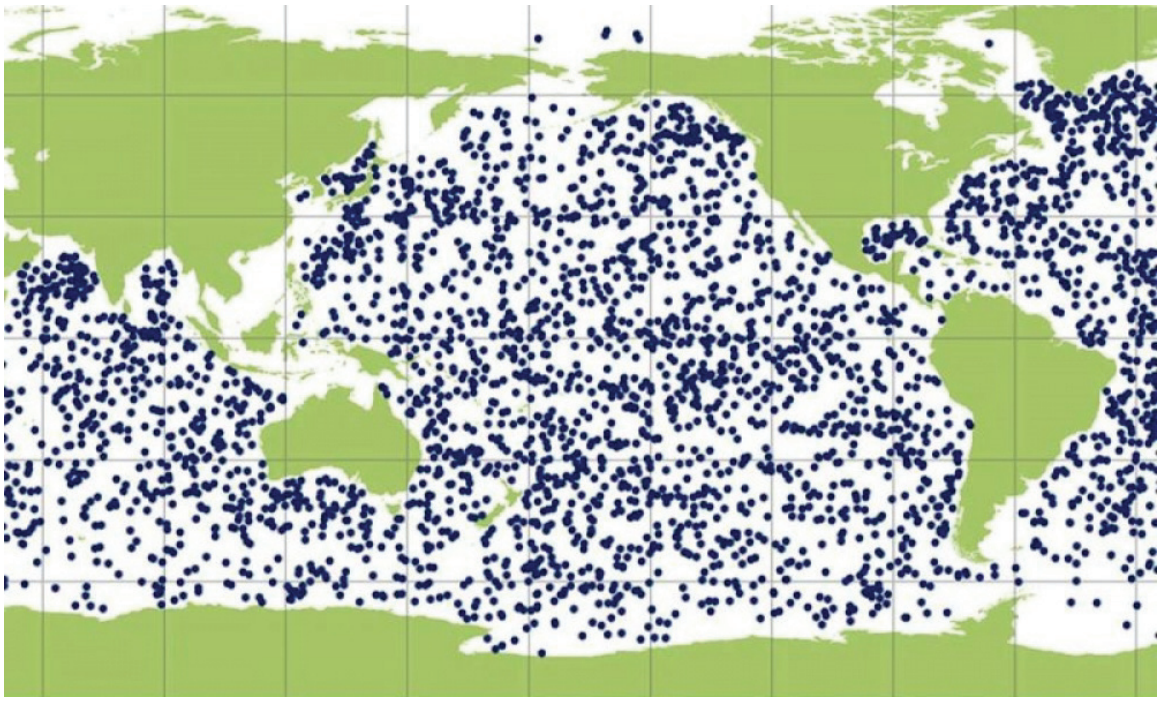


Figure 3: World map displaying 3908 (March 2024) active Argo float locations (https://argo.ucsd.edu/about/status/).

1.4 World Ocean Atlas (WOA)

The WOA offers climatologies for marine physical, chemical, and biological properties. These climatologies are comprised of various datasets collected from multiple sources such as research cruises, moored buoys, and satellite observations. Data interpolation schemes yield a horizontal resolution of 1°x1° (Garcia et al., 2019).

WOA18 is used as a secondary observational data set for evaluation of global ocean models and for comparisons to the Argo data set introduced in 1.3 (BGC-)Argo Floats.

1.5 Objectives

Given the importance of the accuracy of CMIP6 models, several evaluation studies have discussed model biases, with particular interest in changes in model performance from CMIP5 to CMIP6. Most of these evaluation studies (see 1.2) use globally interpolated observational data sets like WOA data for their analyses.

This study aims to evaluate various global ocean models by comparing them to in situ biogeochemical- (BGC-) Argo float observations as well as WOA data. While previous studies have focussed on surface model evaluation (e.g. Laurent et al., 2021) or the comparison of CMIP5 to CMIP6 models (e.g. Fu et al., 2022), this research will present a novel in-depth spatial and temporal evaluation of both surface and subsurface model data including physical and BGC properties, alongside the comparison of data-assimilating models vs. CMIP6.

The motivation for this thesis stems from the widespread utilization of CMIP6 models in crucial studies, often overlooking their inherent limitations, in conjunction with the inquiry into how the performance of data-assimilation models compares to that of CMIP6.

This thesis is guided by the following research objectives: 1) evaluation of whether global ocean models can accurately reproduce North Atlantic Ocean physical and BGC properties, 2) investigate the interrelation between biases in BGC and physical models, and 3) assessment of a CMIP6 model ensemble performance with respect to individual models including data-assimilation models.

This evaluation focusses on the following properties: chlorophyll-a (CHL), nitrate (NO_3), dissolved oxygen (OXY), apparent oxygen utilization (AOU), seawater salinity (SAL), and seawater temperature (TEMP). These properties were chosen to evaluate how important marine BGC processes (e.g., phytoplankton spring bloom) are represented by different global ocean models and whether physical and BGC properties affect each other's performance. Modelled ocean physics and BGC properties affect each other, and information is transferred between the different model components through a coupler (Table 1). Effects of ocean physics on biogeochemistry are for example the distribution of nutrients through advection and mixing, or ocean temperature which affects photosynthesis. BGC properties in turn also affect ocean physics. For example, phytoplankton growth through photosynthesis affects the absorption of solar radiation and the distribution of heat within the ocean.

1.6 Analytical Framework

This thesis adopts a structured approach in presenting its findings from multiple different analyses, with the overall aim to evaluate global ocean models through comparison against observations from (BGC-)Argo and WOA.

Section 2 Data and Methods provides overview over models and data included, as well as any processing techniques. Further, methodologies employed for each individual analysis are introduced, accompanied by definitions of new variables and a detailed explanation of the terminology used. Research questions act as guiding principles, organizing results and discussions into coherent subsections for clarity and understanding.

Beginning with the results, Section 3 lists all findings through introduction and description of the figures. Subsequently, Section 4 discusses and interprets the results cumulatively within the context of the thesis' objectives.

The different analyses included are as follows:

- Annual cycles in the Eutrophic North Atlantic
- Province-averaged Δ-profiles (Model-Argo)
- Spatial comparisons of NO_3, surface and intermediate depth (500-1200 m) biases
- Bias correlation between BGC and physical properties

- Bias splitting into physical and BGC components
- Model ranking through misfit calculation

2 DATA AND METHODS

2.1 Models

Six CMIP6 models were chosen based on previous evaluation studies (e.g. Laurent et al., 2021; Séférian et al., 2020), data availability, and model data accessibility (https://esgf-node.llnl.gov/search/cmip6/). They are CanESM5 from the Canadian Centre for Climate Modelling and Analysis, CMCC-ESM2 from the Centro Euro-Mediterraneo sui Cambiamenti Climatici in Italy, GFDL-ESM4 from the National Oceanic and Atmospheric Administration in the USA, IPSL-CM6A-LR from the Institute Pierre-Simon Laplace in France, MPI-ESM1.2-HR from the Max Planck Institute in Germany, and NorESM2-LM from the Norwegian Meteorological Institute. In addition, the data-assimilating Mercator Ocean International's (MOI) models MOI-BIO4 (for BGC properties) and MOI-GLO12 (for physical properties) are used (Table 1 & Table 2). They utilize observations (e.g., ocean colour chlorophyll, Copernicus Marine Environment Monitoring Service (CMEMS) Operational Sea Surface Temperature (SST) and Ice Analysis (OSTIA) SST/Sea Level Anomaly (SLA)/Sea Ice concentration, World Ocean Atlas (WOA) 13 climatology) for model correction through a reduced-order Kalman Filter based on a SEEK kernel formulation (for detailed information on data-assimilation the reader is referred to Lellouche et al., 2021).

30-year monthly climatological fields (1990-2019) were calculated for CMIP6 models to reduce model-observation differences due to internal and interannual variability on shorter time scales. CMIP6 data was chosen with SSP5-8.5 forcing (worst case scenario) from 2015-2019 to include present day atmospheric carbon dioxide concentration and historical forcing from 1990-2014. The timeframe of 30 years was chosen based on recommendations by the World Meteorological Organization (WMO) for the definition of a climatology (https://www.ncei.noaa.gov/news/new-ocean-regional-climatology-added). For Mercator Ocean International's (MOI's) data-assimilating models, 12-year monthly climatologies will be calculated for data from 2009-2021, as model output before 2009 is unavailable.

Table 1: Overview of models included in evaluation.

Model	Time frame climatology	Horizontal Resolution (Equator)	Physical Model	BGC Model	Coupler
CanESM5	1990-2019	100 km	NEMOv3.4.1	CMOC	CanCPL
CMCC-ESM2	1990-2019	100 km	NEMOv3.6	BFM5.1	CPL v7
GFDL-ESM4	1990-2019	50 km	MOM6	COBALTv2	NEMS
IPSL-CM6A-LR	1990-2019	100 km	NEMOv3.6	PISCESv2	XIOS
MPI-ESM1.2-HR	1990-2019	50 km	MPIOM	HAMOCC6	OASIS3-mct
NorESM2-LM	1990-2019	100 km	MICOM	iHAMOCC	CESM2.1
MOI-BIO4	2009-2021	28 km	NEMOv3.1	MOI-BIO4	-
MOI-GLO12	2009-2021	10 km	NEMOv3.1	-	-

Table 2: Overview of BGC complexity of BGC models included in evaluation.

Model	Nutrients	Phyto- (P) & Zooplankton (Z)	Bacteria
CMOC	**1**: N	**1P, 1Z**	none
BFM5.1	**5**: NH_4, NO_3, PO_4, Si, Fe	**4P**: diatoms, autotrophic nanoflagellates, pico- & large phytoplankton	explicit

Model	Nutrients	Phyto- (P) & Zooplankton (Z)	Bacteria
		4Z: carnivorous & omnivorous mesozooplankton, microzooplankton, heterotrophic nanoflagellates	
COBALTv2	**5**: NH_4, NO_3, PO_4, Si, Fe	**4P**: 3 explicit (small, large, diazotrophs), 1 implicit (diatoms) **3Z**: 1 micro- & 2 mesozooplankton	explicit
PISCESv2	**5**: NH_4, NO_3, PO_4, Si, Fe	**2P**: diatoms, nanophytoplankton **2Z**: micro- & mesozooplankton	implicit
HAMOCC6	**4**: NO_3, PO_4, Si, Fe	**2P**: bulk, cyanobacteria **1Z**	implicit
iHAMOCC	**4**: NO_3, PO_4, Si, Fe	**1P, 1Z**	implicit
MOI-BIO4	**5**: NH_4, NO_3, PO_4, Si, Fe	**2P:** diatoms, nanophytoplankton **2Z**: micro- & mesozooplankton	none

2.2 Observations

The study domain covers the North Atlantic within 0-85°N and 75°W-30°E (see Figure 4). The Argo program serves as the primary observational dataset, supplemented by World Ocean Atlas (WOA) climatologies and satellite CHL data (CHL not available from WOA) for analyses where Argo data are insufficiently dense. WOA and satellite data are additionally employed to compare and validate the different data sets.

(BGC-)Argo float data include profiles taken between 01-01-1999 and 31-12-2021 with increasing numbers of BGC profiles after 2014. Float data were downloaded and processed with the BGC-Argo Matlab Toolbox by Frenzel et al. (2021). Only data with quality flags 1/2/5/8 (very good/good/value changed/estimated value) were used.

WOA18 data is available from the surface to 1500 m depth only and a spatial resolution of 1° by 1° (100 km at the equator) for monthly averages. For Seawater salinity (SAL) and temperature (TEMP) 13-year monthly climatological fields were calculated encompassing the years 2005-2017. BGC properties nitrate (NO_3) and dissolved oxygen (OXY) are available as 118-year monthly climatologies (1900-2017) due to the scarcity of data. Data was downloaded from https://www.ncei.noaa.gov/access/world-ocean-atlas-2018/. Climatologies calculated in this thesis are comprised of different time windows due to data availability and to reduce processing time and storage space.

Satellite CHL data is used in addition to WOA climatologies, as WOA does not offer CHL as one of their properties. Satellite data is NASA's oceandata Aqua Modis L3 mapped data with a spacial resolution of 4.6 km at the equator (https://oceandata.sci.gsfc.nasa.gov/directdataaccess/Level-3%20Mapped/Aqua-MODIS). 18-year monthly climatologies were calculated for data available for the time frame 2003-2021.

2.3 Variables

Model evaluation includes the following properties: Chlorophyll-a (CHL), nitrate (NO_3), dissolved oxygen (OXY), apparent oxygen utilization (AOU), seawater salinity (SAL), and seawater temperature (TEMP).

AOU is defined as:

$$AOU = \text{saturation oxygen } (OXY_{sat}) - \text{observed dissolved oxygen (OXY)} \quad (1)$$

where OXY_{sat} is calculated from the solubility of nitrogen, oxygen and argon in water and seawater as by Weiss (1970). The molar volume of oxygen at standard temperature and pressure is obtained from the NIST website on the thermophysical properties of fluid systems (http://webbook.nist.gov/chemistry/fluid/) using OXY, SAL, and TEMP. AOU is calculated for all models and observational data sets with a Matlab function by Edward T. Peltzer (MBARI, 2007) that requires the inputs OXY, SAL, and TEMP.

AOU can be used to identify how much of the dissolved oxygen is due to biological activity. Positive AOU is defined as biological consumption/utilization of oxygen, negative AOU indicates biological production of oxygen. Dissolved oxygen that is not due to

biological production, and thereby not captured in AOU, comes from the exchange of the surface ocean with the atmosphere.

2.4 BGC-Provinces

The study domain is divided into BGC-provinces adapted from Longhurst (1995) (Figure 13). Every province represents a unique set of environmental conditions. Based on the annual cycle of primary production, eight conceptual models were used to define the BGC-Provinces: polar irradiance-limited production peak, nutrient-limited spring bloom, winter-spring production with nutrient limitation, small-amplitude response to trade wind seasonality, large-amplitude response to monsoon reversal, and various responses to topography and wind stress on continental shelves, including coastal upwelling (Reygondeau et al., 2013).

The division of the domain into provinces was motivated by the assumption that properties within each province share similarity. As a result, the Argo profiles within a given province are theoretically expected to serve as representative samples for that specific region. This partitioning facilitates a more focused analysis.

In the North Atlantic, some provinces describe the shelves. Argo floats do not operate on shelves, which is why these provinces were left out or combined with other provinces. As such, GFST was combined with NWCS and is now called GFST (Figure 14). Further, a region of BPLR was combined with ARCT and is now called ARCT. This region includes waters south of 65°N on the western side of Greenland together with areas to the south and east of Greenland (Figure 14).

Only provinces with “good Argo coverage” are included in the evaluation, which results in seven North Atlantic provinces. A threshold of five profiles per month for every property was set to ensure enough data is available for a province to be labelled “good Argo coverage” (Figure 11). Three such provinces are located in the northern North Atlantic (ARCT, SARC, NADR), 3 in the southern North Atlantic (NASW, NATR, WTRA), and GFST is in the transition zone (Figure 4).

Each float profile was sorted into BGC provinces as some floats cover more than one province. There are different numbers of profiles for each property in each province due to different types of floats and quality control (Table 4).

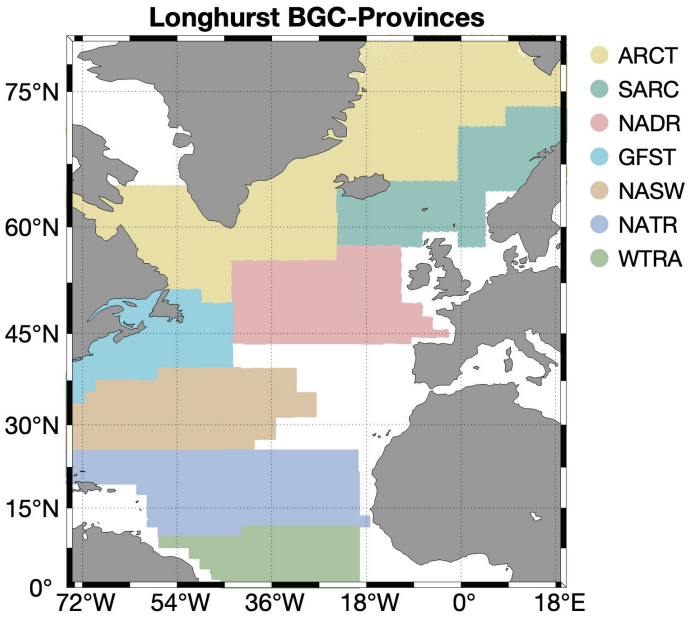


Figure 4: Adjusted BGC-provinces with "good Argo coverage" in the North Atlantic. ARCT: Arctic, SARC: Southern Arctic, NADR: North Atlantic Drift, GFST: Gulf Stream, NASW: North Atlantic Subtropical Gyre, NATR: North Atlantic Tropical Gyre, WTRA: Western tropical Atlantic.

2.4 Interpolation Methods

To reduce processing time and to facilitate comparison amongst models and observations, all 6 CMIP6 models and WOA data were interpolated onto the same spatial and depth grid as MOI-BIO4 (1/4° resolution, 40 depth levels from 0-2000m with higher depth resolution near the surface). MOI-GLO12 was interpolated onto MOI-BIO4's spatial grid only, as depth resolution is the same for these models.

All Argo profiles were then interpolated onto the same MOI-BIO4 depth vector. For each Argo profile, the closest model location was identified. In some provinces, Argo profiles are spread unevenly throughout the province (Figure 15). To reduce the effects that could be introduced in a comparison of province-wide averaged observations with model data,

all analyses in this evaluation compare Argo profile data vs. model data at the same location only, unless specifically stated otherwise.

For profiles near a coast, model resolution is too coarse to allow for accurate comparison. These profiles were excluded.

To visualize differences between Argo float observations and models, “average Δ-profiles” were calculated: Argo profile data were subtracted from model data at the closest location to the profile, for each depth, and for the same month. For example, an Argo profile that was recorded in March is subtracted from the model March climatological field. Then, an average Δ-profile per province was calculated.

2.5 Standard Deviation (STD)

STDs are computed for each depth level using Argo data, serving as an estimate of the natural variability within the observations. In this thesis, *natural variability*, is used as a terminology for all variability, including anthropogenic signals such as ocean warming due to elevated carbon emission, and any variability caused by events and processes on varying timescales (e.g. the diurnal cycle of zooplankton affects nutrient concentrations on timescales of hours, while El Niño (interannual variability) happens on the timescale of 2-8 years and has effects on ocean temperature and nutrient distribution through changes in upwelling patterns amongst other effects.). It's important to note that the *actual natural* variability may exceed these estimates, particularly due to the sparsity of Argo data and the spatial clustered distribution of float profiles within a province. However, given the definition of BGC provinces, properties are estimated to be similar within one province. For this reason, the standard deviation will be used here as an estimate for the natural variability and be referred to as such.

The process involved calculating STDs by initially determining the STD from all profiles within a specific province for the same month at each depth level (e.g., STD for each depth from all March profiles in NADR). Subsequently, an average STD for each depth level within a province was computed based on the STDs obtained in the first step.

For Annual Cycle analyses, an average STD was calculated from all measurements between 0-100m for each month in each province.

The STD is defined as:

$$\sigma_{Argo} = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}} \quad (2)$$

With x_i representing each individual measurement at a certain depth level, $\overline{x}$ stands for the mean of all data points at that depth level, and n is the total number of data points at that depth level.

2.6 Misfit

Misfit between models and Argo observations were calculated for each property, province, and depth. Misfit is defined following a method by Rickard et al. (2016) and Laurent et al. (2021) who used this approach to evaluate CMIP6 predecessors (CMIP5).

$$misfit = \overline{abs(BIAS)} + \overline{rms(BIAS)} \quad (3)$$

BIAS is defined as the average Δ (model-Argo) for each property, within each province and for each depth level at every month of the year (e.g. the average of 30 January NO_3 profiles taken in ARCT for each depth level).

Misfits were calculated from averages of abs(BIAS) and rms(BIAS): first annual averages were calculated for abs(BIAS) and rms(BIAS) (e.g. average over all January averages etc.), then those were further averaged over a) the euphotic zone (0-110m, 23/40 depth levels) and over b) below the euphotic zone (130-2000m, 17/40 depth levels).

These misfit calculations are used as a foundation for model performance ranking.

2.7 Bias splitting

A method analogue to the TA* approach (Koeve et al., 2014; Feely et al., 2002; Sabine et al., 2002; Chung et al., 2003) is used to split NO_3 biases into two components: A preformed part (NO_3^0) that is affected by ocean physics (e.g. circulation, freshwater flow) and a remaining part (NO_3^{bio}) due to ocean biogeochemistry (i.e. remineralization).

$$\Delta NO_3 = \Delta NO_3^0 + \Delta NO_3^{bio} \quad (4)$$

The preformed part is constructed from the concentration of NO_3 the water parcel had when it was last in contact with the atmosphere. All water masses are mixtures of various components for which both the component concentrations and the component ratios are unknown in the ocean interior. ΔNO_3^0 can be determined empirically by multilinear regression of ΔNO_3 with ΔSAL, ΔTEMP, and ΔNO.

NO is a conservative tracer as by definition of Broecker (1974). NO is based on the assumption that O_2 and NO_3 can be combined in a way that alterations by respiration are zero. Broecker (1974) suggests that for one mole of O_2, roughly 1/9 mole of nitrogen is released. This ratio has since been updated to $r_{O2:NO3} = 10.625$ (Anderson and Sarmiento, 1994; Kanamori and Ikegami, 1982).

$$\Delta NO = \Delta O_2 + r_{O_2:NO_3} * \Delta NO_3 \quad (5)$$

In a first step, multilinear regression of the surface concentrations is used to determine the regression coefficients (α_{SAL}, α_{TEMp}, α_{NO}) (Eq. 6). In a second step, the calculated coefficients are applied to ΔSAL, ΔTEMP, and ΔNO for calculations of the ocean interior (Eq. 7).

$$\Delta NO_3^{surf}(x,y) = \alpha_0 + \alpha_{SAL}\Delta SAL^{surf}(x,y) + \alpha_{TEMP}\Delta TEMP^{surf}(x,y) + \alpha_{NO}\Delta NO^{surf}(x,y) \quad (6)$$

$$\Delta NO_3^0(x,y,z) = \alpha_0 + \alpha_{SAL}\Delta SAL(x,y,z) + \alpha_{TEMP}\Delta TEMP(x,y,z) + \alpha_{NO}\Delta NO(x,y,z) \quad (7)$$

With this method, biases in modelled NO_3 will be investigated with regards to the contributions of model physics and model biogeochemistry.

3 RESULTS

3.1 Introduction

In this section, cumulative findings from the various analyses listed in 1.6 Analytical Framework will be presented. Each subsection is dedicated to addressing a concrete inquiry, guided by a subsection-specific research question. Given that each analysis is conducted individually for every province, figures illustrating noteworthy trends will be provided only for provinces displaying significant highlights, as many provinces often exhibit similar patterns. All omitted figures can be referenced in the appendix.

3.2 How do float observations compare to WOA observations?

Due to the spatial distribution of the Argo data set (clustered profiles for some properties, see Figure 15), it is advisable to validate the Argo dataset by comparing it to another data set, such as WOA. While WOA has been proven to be a successful tool for model evaluation (e.g. Séférian et al., 2020), the Argo dataset represents a newer approach and requires such validation.

Average annual cycles in the productive northern North Atlantic provinces are used to investigate whether province-wide float coverage is good enough to display expected annual cycles in the upper 110 m and whether these align with observations from the WOA. Annual cycles for the southern part of the study domain are expected to be less pronounced and are therefore left out of the comparison. ±1 standard deviation (σ) of Argo observations indicates estimated natural variability of the float observations (visualized by grey shading in figures).

Annual cycles of WOA data and Argo agree well, but an approximately 10 µmol/kg AOU offset is identified with Argo AOU concentrations being higher. WOA data lies mostly within ±1 standard deviation of Argo observations (Figure 5).

Average Δ-profiles (WOA-Argo) are used to validate full depth (0-2000 m) Argo profiles. For CHL, satellite observations were added to depth profiles to validate Argo data. Throughout the water column (0-2000 m), WOA data lies within the natural variability of

float observations (Figure 6) and satellite CHL compares well with Argo data at the surface in northern provinces. In southern provinces, ±1 standard deviation of satellite observations overlap with ±1 standard deviation of Argo data (Figure 6).

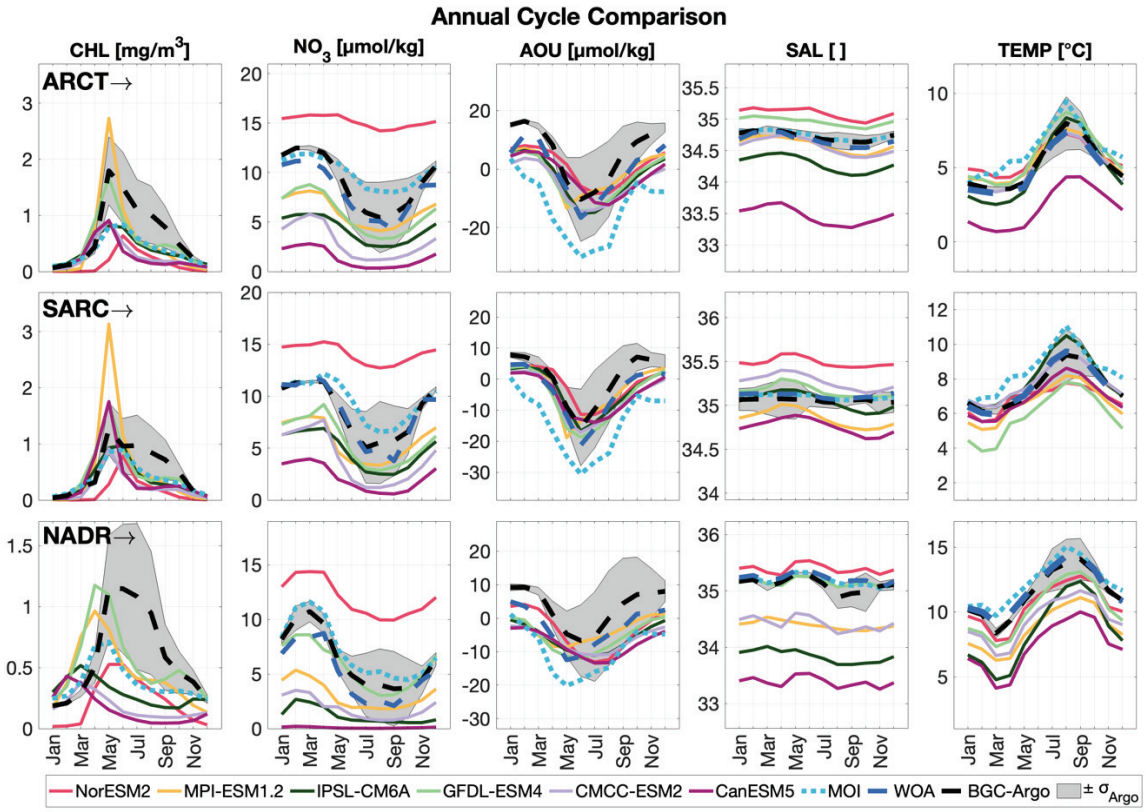


Figure 5: 0-110m averaged annual cycles of northern North Atlantic provinces ARCT (top panels), SARC (middle panels), and NADR (bottom panels) for properties CHL, NO_3, AOU, SAL & TEMP. Argo data in dashed black, WOA in dashed blue, MOI in light blue dotted line, and CMIP6 models in remaining solid colours. WOA, MOI, and CMIP6 data from locations of Argo profiles only. Grey shading displays ±1 standard deviation of Argo data.

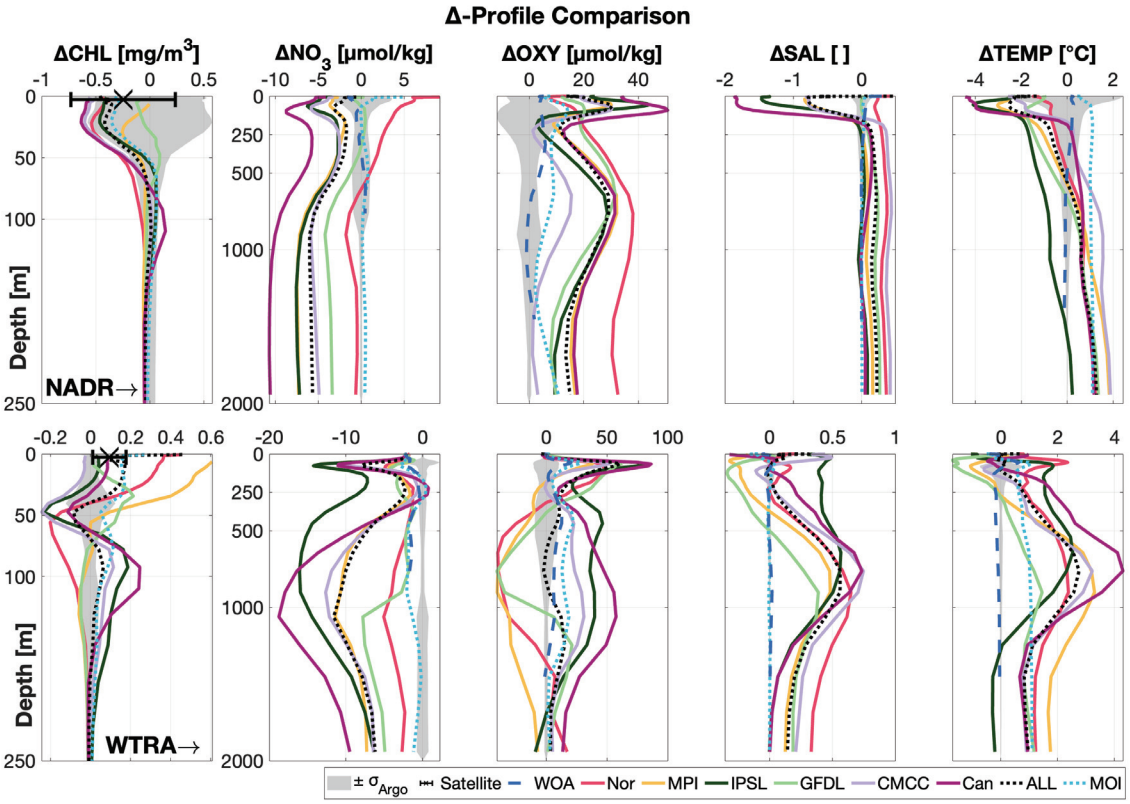


Figure 6: Average delta (model-Argo) depth profiles for CHL, NO_3, OXY, SAL & TEMP for provinces NADR (northern North Atlantic) and WTRA (southern North Atlantic). Black x with ±1 standard deviation indicates satellite CHL, Argo data in dashed black, WOA in dashed blue, MOI in light blue dotted line, CMIP6 models in remaining solid colours, and the black dotted line is an ensemble of all CMIP6 models (ALL). WOA, MOI, and CMIP6 data from locations of Argo profiles only. Grey shading displays ±1 standard deviation of Argo data. CHL profiles shown to 250m depth, other properties to 2000m.

3.3 Do global ocean models compare well with observations?

3.3.1 Surface Comparison

Annually averaged surface maps of WOA, CMIP6 and MOI data are shown for comparison of modelled surface concentrations with observations. Largest differences amongst models and between models and WOA are found for NO_3 (Figure 7, Figure 8, Figure 18, Figure 19).

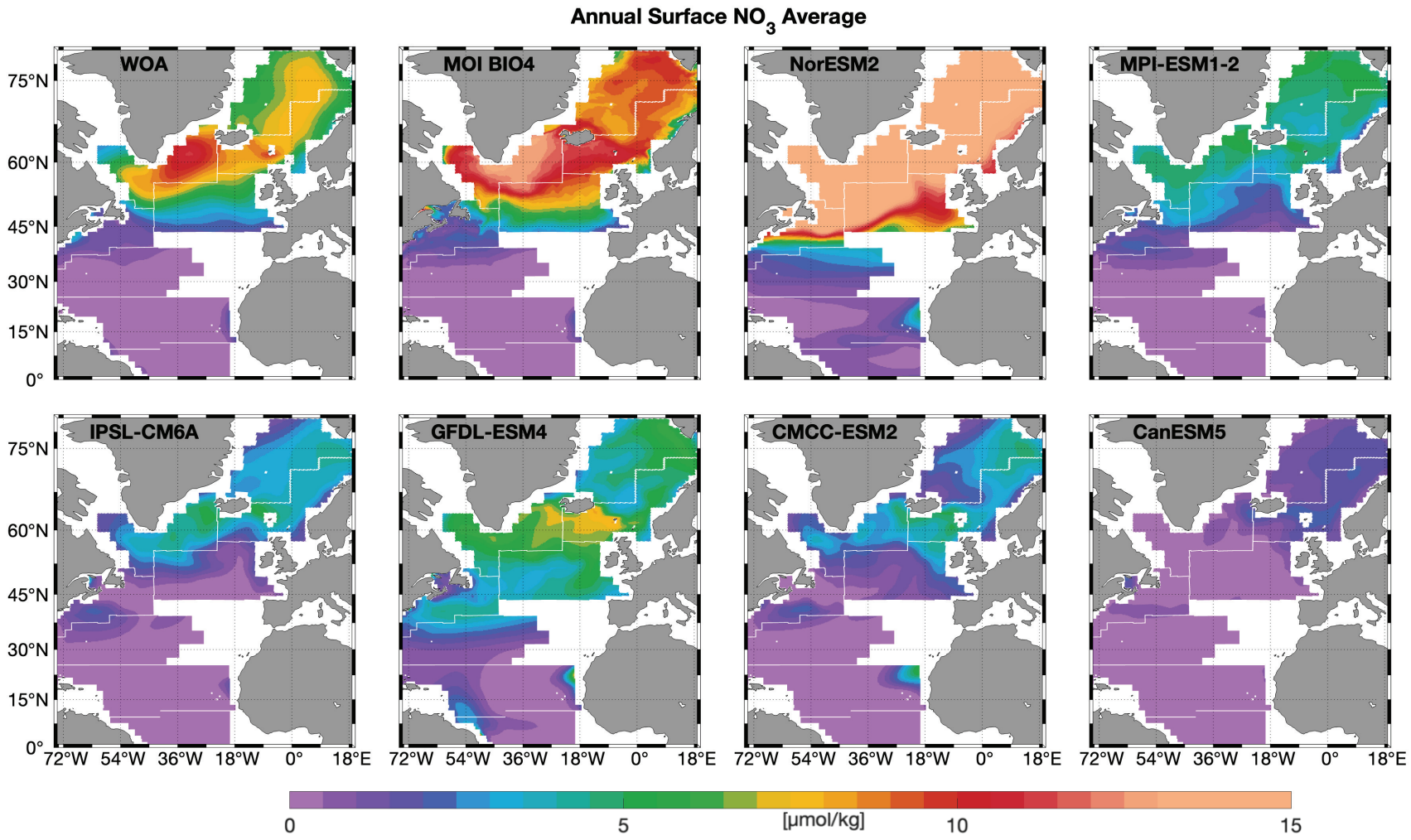


Figure 7: Annual surface NO_3 [µmol/kg] average. Top row left panel shows WOA annual average from 1900-2017. Top row second from the left is data-assimilating model MOI from 2009-2021. Remaining panels are CMIP6 models, averaged over 1990-2019. Surface refers to an average over 0-6m.

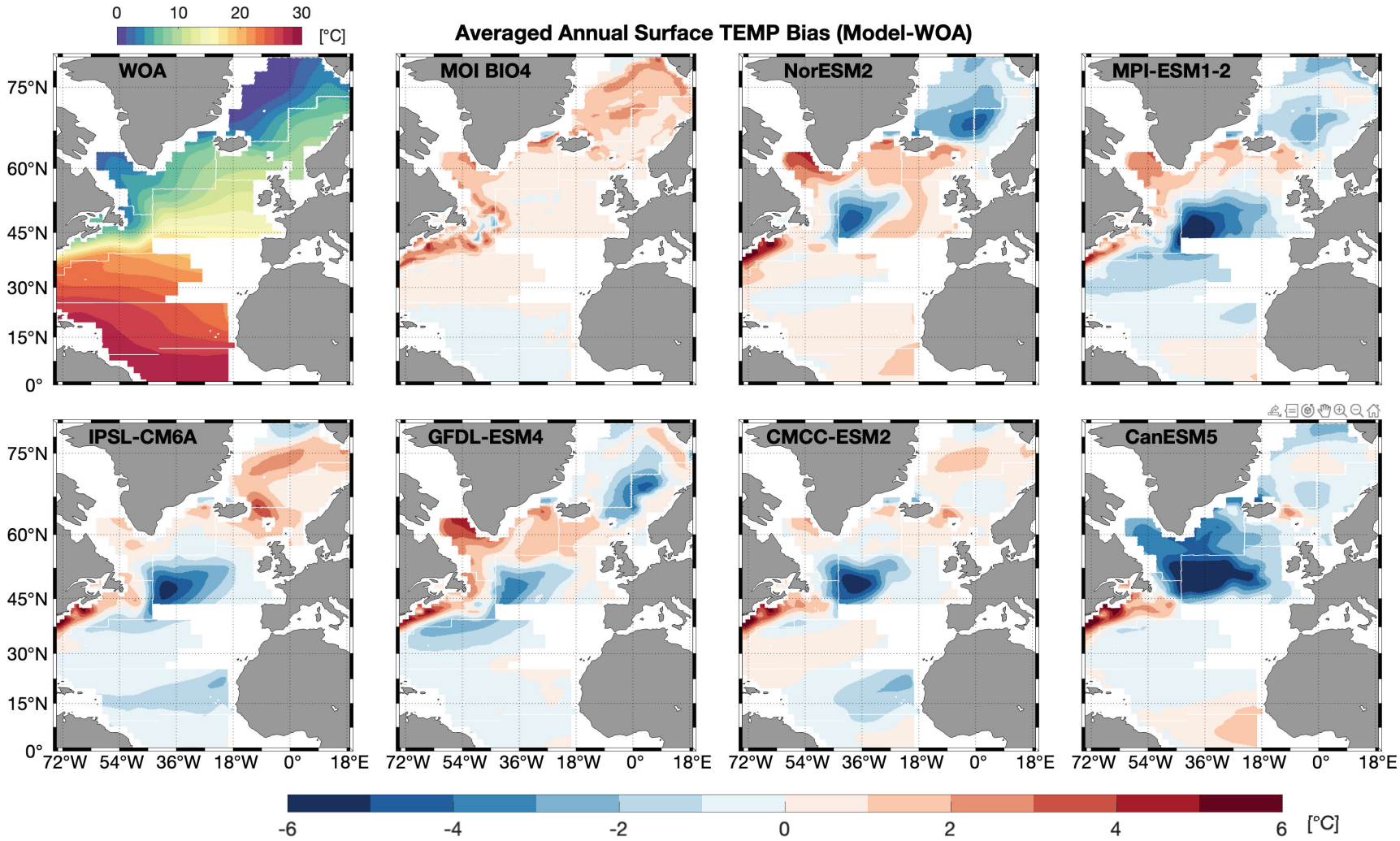


Figure 8: Annual average surface (0-6m) TEMP bias (Model-WOA) maps. Top row left panel shows WOA annual average from 1900-2017. Top row second from the left is data-assimilating model MOI from 2009-2021. Remaining panels are CMIP6 models, averaged over 1990-2019.

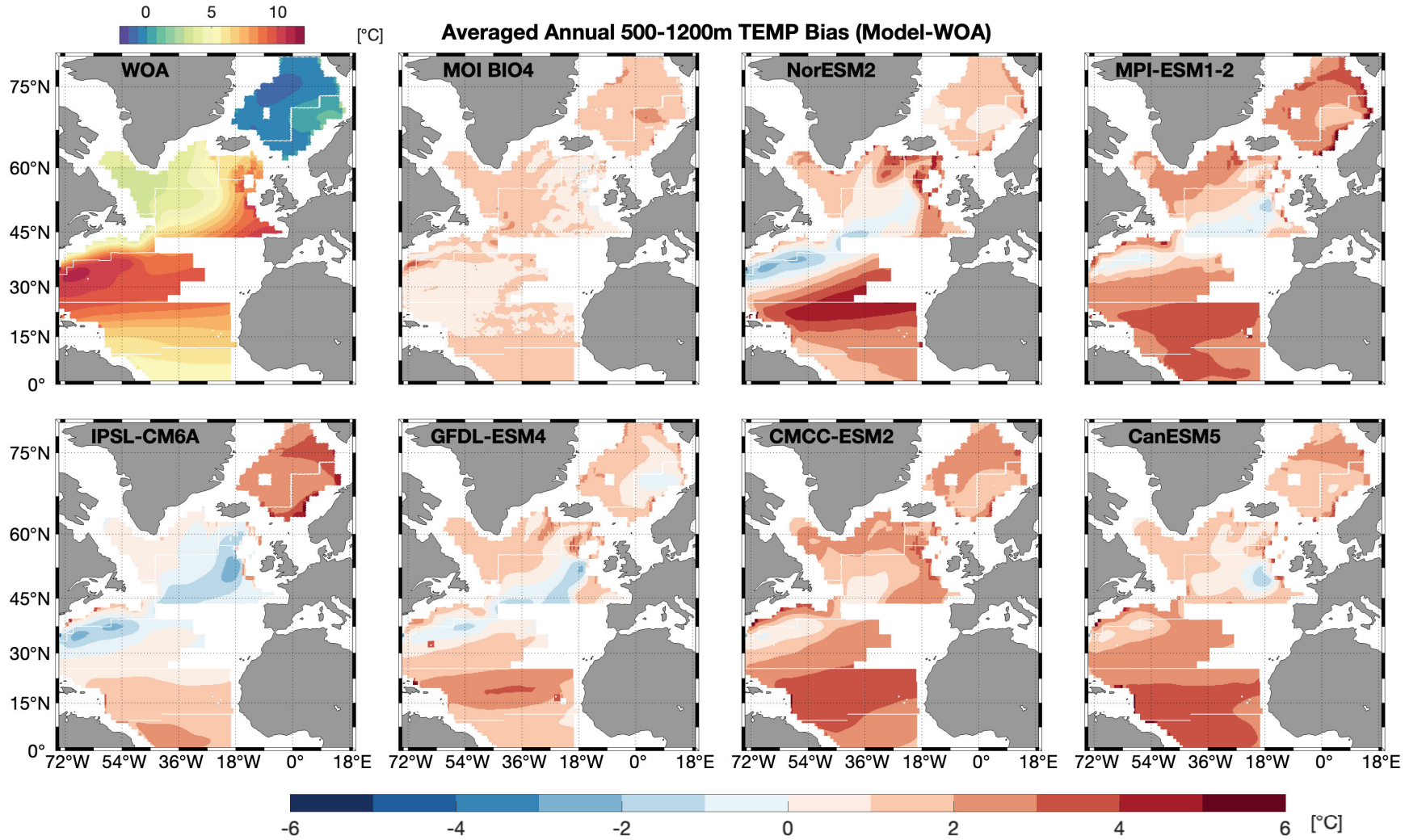


Figure 9: Annual average (500-1200) TEMP bias (Model-WOA) maps. Top row left panel shows WOA annual average from 1900-2017. Top row second from the left is data-assimilating model MOI from 2009-2021. Remaining panels are CMIP6 models, averaged over 1990-2019.

Annually averaged surface NO_3 concentration varies in concentration and spatial distribution between WOA and CMIP6 models: There is no consensus amongst CMIP6 models in terms of over- or underestimations (e.g. NorESM2 overestimates surface NO_3 north of 45°N, while CanESM5 underestimates surface NO_3 in the entire domain). Out of six CMIP6 models, five models underestimate surface NO_3 concentrations north of 45°N. NorESM2, GFDL-ESM4 & CMCC-ESM2 overestimate surface NO_3 concentrations in the upwelling region off of the western coast of Mauretania (around 20°N, 36-18°W) (Figure 7).

Data-assimilating MOI compares best with WOA data, but an overall overestimation of surface NO_3 of 2-3 µmol/kg in the northern North Atlantic is visible (Figure 7).

Annual average surface concentrations of OXY, SAL and TEMP are overall better captured by CMIP6 models and MOI models than NO_3 (Figure 8, Figure 18, Figure 19). Local biases are found in the Gulf Stream region, where a TEMP overestimation close to the shelf and coast turns into an underestimation towards the east. SAL is overall overestimated and OXY underestimated in this area. Additionally, TEMP is strongly underestimated at around 50°N, where the North Atlantic Current (NAC) transports warm salty waters to the North. SAL is underestimated by all models in this region and OXY is strongly overestimated, but the spatial extent of OXY biases varies with models. As such, CanESM5 overestimates surface OXY in the entire Labrador Sea and south of Greenland, while NorESM2 underestimates surface OXY in the Labrador Sea, but shows a smaller patch of overestimations south of Greenland at 50°N. In the Greenland and Norwegian Sea, modelled TEMP and SAL do not agree on a direction of bias. MOI and IPSL-CM6A overestimate TEMP here, while the remaining models tend to underestimate surface TEMP. SAL biases are weak in this area, while OXY tends to be overestimates with varying spatial extend and magnitude of the bias.

MOI surface properties agree best with WOA observations, which is to be expected given the assimilation of data (WOA13 amongst others).

3.3.2 Annual Cycles in the northern North Atlantic

Seasonality is most pronounced in the northern half of the study domain (ARCT, SARC, NADR). As such, Figure 5 displays average annual cycles are displayed for these provinces only for different properties.

In northern hemispheric winter (Jan-Mar), water temperatures are lowest throughout the year in all 3 provinces. Starting in March, CHL concentrations increase together with a drawdown of nitrate until April/May when CHL peaks. Nitrate is lowest approximately three months after. Coexistent with the increase in CHL is a decrease in AOU, meaning that oxygen is being produced. CHL concentrations decline one month after the spring peak, slightly rise again in August/September (fall bloom) and reach lowest values in December. AOU increases starting in May/June, about one month after the CHL drawdown. Oxygen is being utilized during this second half of the year. A second decrease in AOU representing the fall bloom, is absent in the observations. Nitrate starts to increase in around September until the end of the year. Highest temperatures are reached in August and start to decrease thereafter until Dec.

All six CMIP6 models follow above described overall seasonal pattern. Largest differences are visible for CHL, for which the drawdown after the peak in May is too rapid in most models. Timing of the spring peak is biased for NorESM2 throughout the northern provinces and for other models in NADR. Modelled CHL peak magnitudes are biased with an overall underestimation. A fall CHL peak is visible for GFDL-ESM4, CanESM5 & MPI-ESM1.2, but neither magnitude nor timing align well with the observations. CMIP6 NO_3 and TEMP annual cycles are overall underestimated and modelled SAL cycles do not agree on a direction of bias (Figure 5).

Data-assimilating MOI models agree better with Argo observations than CMIP6 models. However, MOI models underestimate AOU by up to 20µmol/kg. Magnitudes of the CHL spring bloom are also underestimated, but the timing aligns well with the observations. NO_3 and TEMP are slightly overestimated by MOI models (Figure 5).

3.3.3 Physical properties in the North Atlantic

This section describes biases found for the properties TEMP and SAL which are affected by advection, mixing, and other physical processes. Apart from previously discussed surface biases found in the Greenland and Norwegian Sea, as well as the Gulf Stream region, largest physical biases are located at the top 150 m and between 500-1200 m depth (Figure 6). While mixed layer biases are more common in the northern North Atlantic provinces where both SAL and TEMP are underestimated by CMIP6, intermediate depth biases occur as strong overestimations and are pronounced in the southern half of the study domain. Physical biases at depth are also prominent in the eutrophic North Atlantic, as consistent overestimations of TEMP and SAL with depth (>300 m), rather than confined to 500-1200 m.

Although, surface maps (0-6 m) of TEMP and SAL indicate good agreement between models and observations, average bias maps of intermediate depth (500-1200 m) visualize those biases found in the depth profiles more clearly (Figure 9). Intermediate depth TEMP is overestimated by CMIP6 in the Greenland and Norwegian Sea where coldest TEMP are expected, and in the centre of the North Atlantic Gyre. The Gulf Stream region, which has the warmest TEMP in the observations, is however underestimated by all CMIP6 models.

SAL biases at intermediate depth are less pronounced than those of TEMP. SAL is overestimated in the centre of the North Atlantic Gyre where SAL is highest. Gulf Stream SAL is slightly overestimated by CMIP6. CanESM5 underestimates intermediate SAL in the Greenland and Norwegian Sea.

MOI intermediate TEMP and SAL agree significantly better with WOA observations than CMIP6 does: While most pronounced biases between MOI and WOA range from -1 to 3°C for TEMP and -1 to 2 for SAL, CMIP6 models display discrepancies as pronounced as -3 to 5°C for NorESM2 TEMP and -3 to 3 for IPSL-CM6A SAL.

3.4 Are global ocean models in good agreement within the natural variability of the observations?

Overall, CMIP6 modelled properties lie outside of the estimated natural variability (see 2.5 Standard Deviation (STD)) of Argo observations, while MOI data agree better with such natural variability (Figure 5, Figure 6).

CMIP6 performance shown in averaged Δ-profiles (model-Argo) is similar across properties amongst provinces in the northern part of the domain (ARCT, SARC & NADR), as are Δ-profiles amongst provinces in the southern provinces (Figure 17). NADR and WTRA were chosen as example provinces for the eutrophic northern North Atlantic and the oligotrophic South (Figure 6).

For all five properties, CMIP6-Δs lie outside of the natural variability of the observations at multiple depths: Overall model performance is best for CHL in NADR. CHL is the only property for which up to three model-Δs (MPI-ESM1.2, IPSL-CM6A & GFDL-ESM4) lie within the natural variability of the observations in the entire water column (up to 250 m for CHL). CHL biases are visible for the remaining CMIP6 models and are strongest in the upper 50 m. At these depths, CHL is being underestimated (overestimated) by to 0.6 mg/m^3 in the northern provinces (oligotrophic provinces).

Large biases for NO_3, OXY, SAL & TEMP are found in the upper 150 m and at 500-1500 m. NO_3 is underestimated by all CMIP6 models but NorESM2 in the upper 500 m in the eutrophic North Atlantic. Underestimations reach -18 µmol/kg for CanESM5 at depths ~1000 m. OXY is overall overestimated by all CMIP6 models in the northern province but underestimated by three out of six CMIP6 models at depths >300 m in the southern provinces. Strongest OXY overestimations occur at 50-150 m for OXY (up to 80 µmol/kg for CanESM5). SAL and TEMP biases indicate underestimations in the upper 150 m in the eutrophic northern provinces and strong overestimations at intermediate depths in the oligotrophic South.

GFST averaged Δ-profiles (Figure 17) are similar to those in the northern provinces, but CMIP6 performance is better at the surface (0-100 m), especially for OXY and TEMP.

MOI model-Δs are in significantly better agreement with the natural variability of Argo observations than CMIP6 models. While SAL Δ-profiles align perfectly with the natural variability, NO_3 profiles in the southern provinces are underestimated and OXY& TEMP profiles are overall overestimated. MOI CHL agrees well with the natural variability of Argo observations in the northern provinces but is overestimated in the South (Figure 6).

Over 80% of all CMIP6 and around 40% of MOI average model Δ-profiles lie outside of the natural variability of the observations (Figure 5, Figure 6).

3.5 Are BGC biases linked to biases in physical properties?

The similarity of bias patterns in Δ-profiles across properties indicates linkages between biases of different properties. Especially at depth 500-1200 m, large biases were identified in southern North Atlantic provinces. Figure 10 is used to investigate possible correlations: TEMP-Δs at 500-1200m were plotted against NO_3-Δs at these depths and Pearson's correlation coefficients were calculated for each model (shown in legend). Regression lines were added for visualization purposes.

In the entire southern half of the study domain, both CMIP6 model biases and MOI model biases for TEMP and NO_3 are anticorrelated at intermediate depths. Variability around these anticorrelations varies amongst provinces. In NATR anticorrelations range from -0.44 (MOI) to -0.93 (GFDL-ESM4, MPI-ESM1.2) (Figure 10).

Biases in northern North Atlantic provinces NADR and ARCT are not anticorrelated, however in SARC anticorrelations range from -0.54 (MOI) to -0.76 (GFDL-ESM4) (Figure 21).

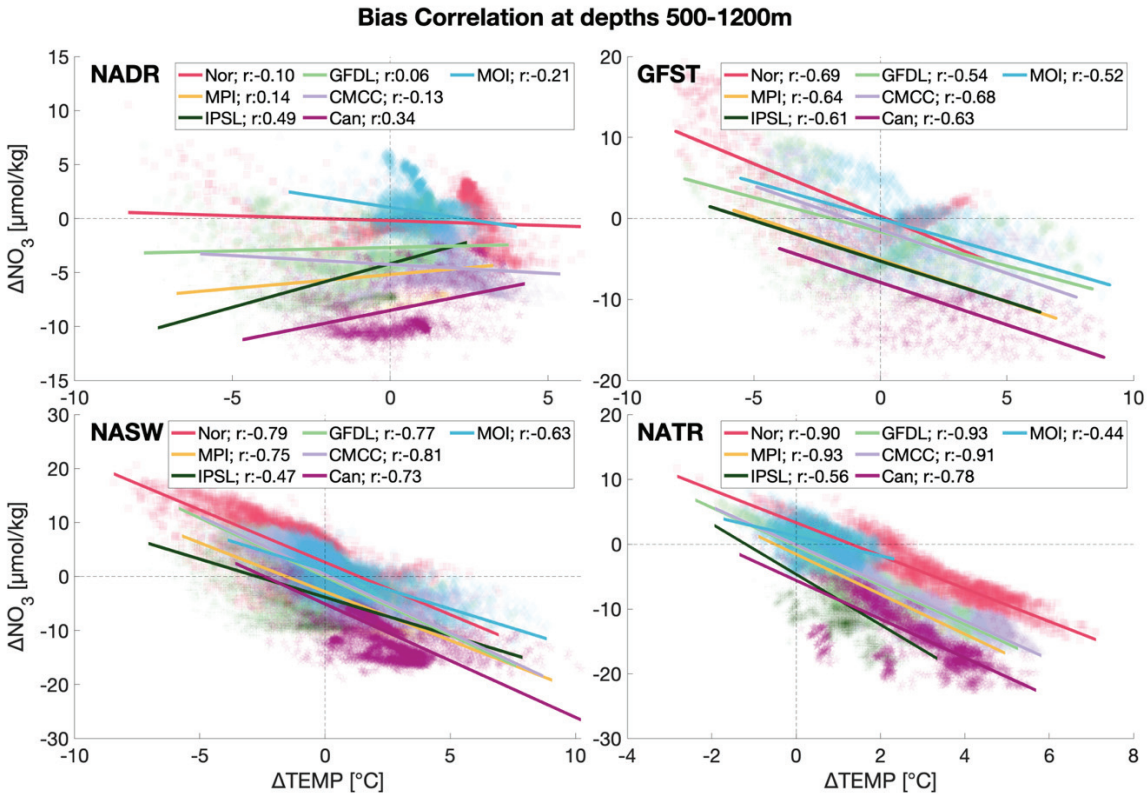


Figure 10: TEMP biases vs. NO_3 biases at 500-1200m for provinces NADR, GFST, NASW & NATR. Coloured lines indicate regression for each model, Pearson's correlation coefficients r are given in legend. Negative r indicates anticorrelation.

By using an adaptation of the TA* approach (see 2.7 Bias splitting; Eq. 4), NO_3 biases can be split into a preformed part affected by model physics and a biogeochemically driven remaining bias. Figure 11 displays depth profiles for NO_3 biases, their preformed part and their BGC bias component. In northern North Atlantic provinces (e.g. ARCT in Figure 11) the preformed part of the NO_3 bias is strong and follows the same pattern as the overall NO_3 bias. BGC biases (NO_3^{bio}) are small at the surface and increase at depth >100 m. While MPI-ESM1.2 NO_3^0 and NO_3^{bio} bias components add up to a larger overall ΔNO_3, BGC driven biases of NorESM5 balance out part of the preformed NO_3 bias.

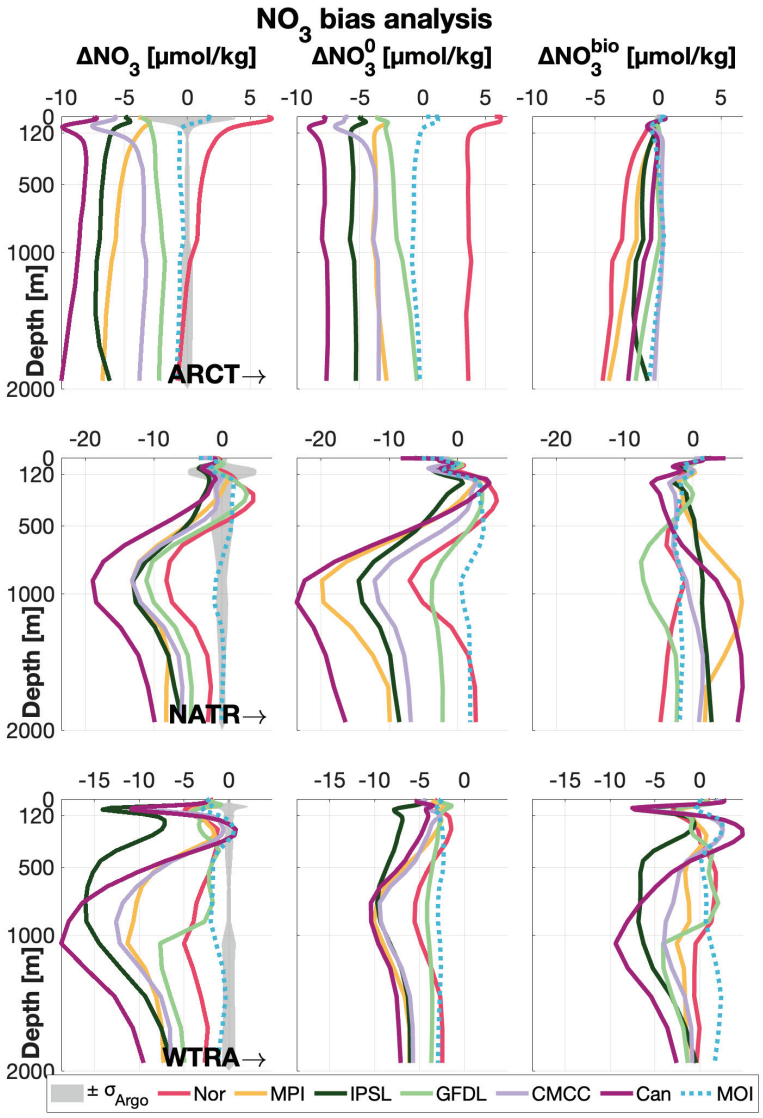


Figure 11: NO_3 bias analysis, splitting the NO_3 bias (left panels) into a component based on ocean physics (preformed NO_3 bias: NO_3^0, vertical middle panels) and a bias component due to BGC processes (NO_3^{bio}, right panels) for three example provinces ARCT (top panels), NATR (horizontal middle panels), and WTRA (bottom panels). Solid coloured lines refer to CMIP6 model biases, dotted blue line indicates MOI data-assimilation model.

A similar effect can be observed in southern North Atlantic province NATR: GFDL-ESM4 displays the pattern of bias addition, while NO_3^{bio} biases are reducing the preformed NO_3 bias for now both NorESM5 and MPI-ESM1.2. Additionally, as for northern provinces, preformed NO_3 bias profiles follow the pattern of overall NO_3 biases in ARCT. However, differently from northern provinces, both bias components are strong in the top 100 m and of opposite sign (underestimation of preformed NO_3 and overestimation of BGC NO_3).

In WTRA, both bias components are equally large and are overall underestimations. In the top 100 m, NO_3^{bio} biases follow the pattern of ΔNO_3 more closely then preformed biases.

3.6 Do individual global ocean models outperform a CMIP6 ensemble?

To rank model performance, misfits between modelled data and Argo observations were calculated for each property in each province (see 2.6 Misfit).

The model with the lowest misfit ranks 1st and the model with the highest misfit has the highest rank out of 8 (six CMIP6 models, CMIP6 ensemble mean & MOI models as MOI).

Model ranks are variable amongst properties and provinces: For example, model misfit for NorESM2 (Nor) is high in ARCT and SARC for CHL, but relatively low in these provinces for NO_3 (Figure 12, top two panels). While CHL in WTRA is best modelled by IPSL-CM6A, this model's NO_3 has largest misfits in WTRA. The ensemble mean performance is as variable as individual model performance.

AOU model performance can be independent of TEMP, SAL and OXY ranking, as biases can "cancel each other out". For example, GFDL-ESM4 modelled OXY and TEMP have largest misfits in SARC, but GFDL-ESM4 ranks first for AOU in that province.

An overall average rank was calculated, and models were ranked again based on this average rank (Figure 12, bottom panel).

MOI data-assimilating models have the overall lowest average misfit. However, for some properties and provinces, individual CMIP6 models can outperform MOI models.

Within the top 110 m, MPI-ESM1.2 ranks second. Below 110 m and over the entire 2000 m profiles, GFDL-ESM4 outperforms other CMIP6 models (Figure 12, bottom panel).

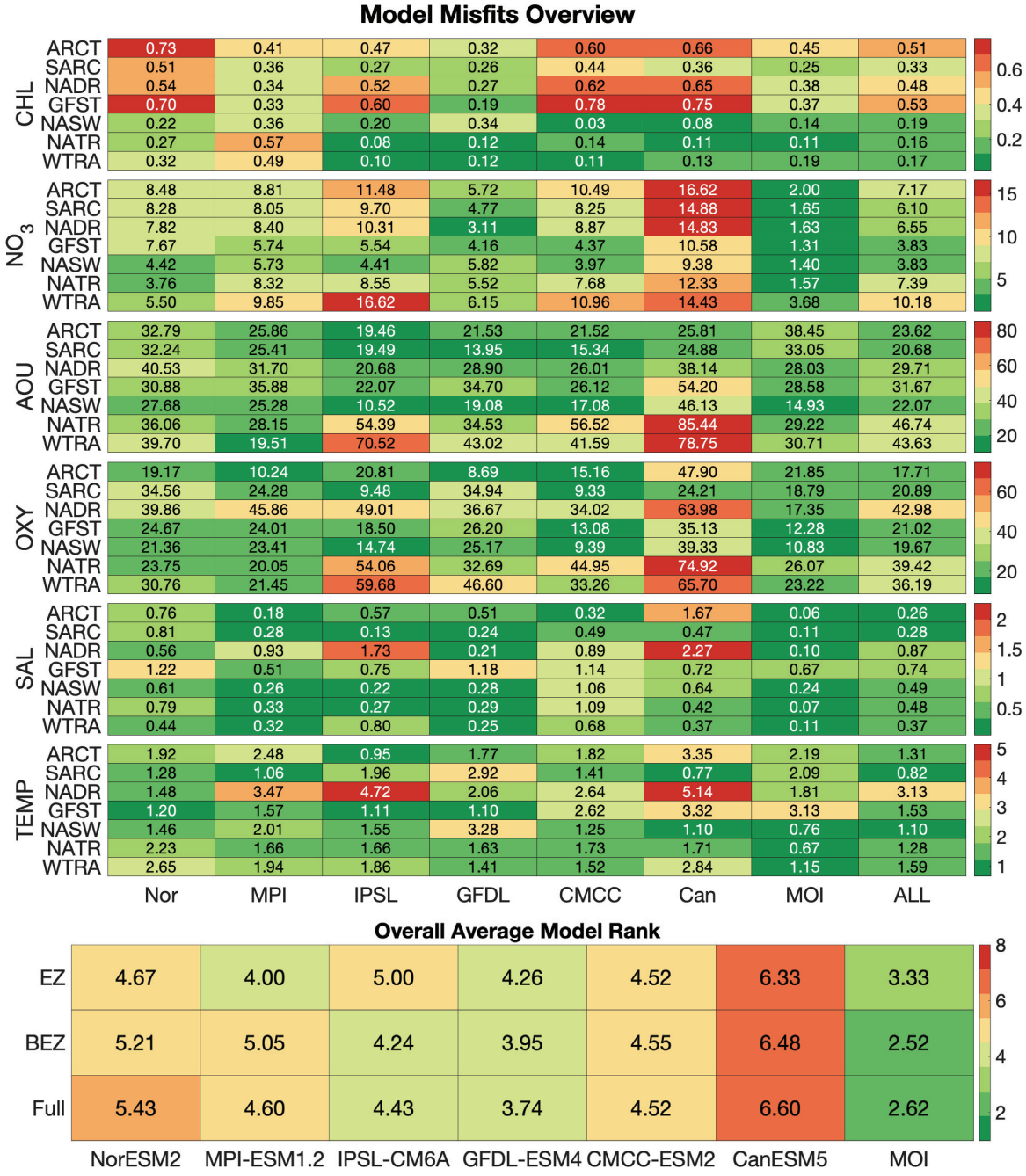


Figure 12: Top: Model misfits for each province (rows) for each property (panels). ALL denotes CMIP6 ensemble mean. Bottom: Average model rank based on lowest misfit for Euphotic Zone (EZ) (0-100 m), below Euphotic Zone (BEZ) (130-2000 m), and full profile (0-2000 m). Lowest misfit/rank indicates best model. ALL is excluded given that individual models outperform ensemble mean.

4 DISCUSSION

4.1. Evaluating whether global ocean models can reproduce an accurate state of North Atlantic properties.

Biases between modelled data and Argo observations were found for all models for all variables in all provinces (Figure 5 ,Figure 6, Figure 7, Figure 8, Figure 10). While it is expected that global ocean models may contain some degree of bias, here evaluated CMIP6 models are presumed to be capable of effectively addressing research inquiries regarding climate change, among other applications. For a model to be deemed suitable for future climate projections, it must minimize biases to accurately represent the current state of ocean properties. However, this alone might not be sufficient for accurate climate projection as the ability to model the response to anthropogenic forcing is different from the ability to reproduce a realistic background state of properties.

The biases found in this thesis indicate a systematic underrepresentation of nitrate and a systematic overestimation of euphotic zone OXY (Figure 6, Figure 17), next to multiple misrepresentations of ocean physics affecting the location of major ocean currents (Figure 8, Figure 9, Figure 10, Figure 19). These findings are similar to Laurent et al. (2021) who report underestimated CHL, variable NO_3 amongst models, and biased circulation patterns for CMIP5 models (CMIP6 predecessors). Séférian et al. (2020) found a greater diversity of BGC models used for CMIP6 than CMIP5 which could explain little consensus amongst CMIP6 model biases reported in this study (see 4.3 Assessment of a CMIP6 model ensemble performance with respect to individual models including data-assimilation models.), and persisting model-data errors as for example biased OXY concentrations at 150 m in the tropical Atlantic, which was also reported here (WTRA in Figure 6).

These results indicate that currently global ocean models have difficulties with the representation of North Atlantic properties, particular ones being reviewed in the following.

4.1.1 Underestimated phytoplankton blooms in the northern North Atlantic

While modelled surface (0-6 m) OXY, TEMP & SAL compare well to WOA data, there are significant discrepancies between CMIP6 and WOA and large variations amongst CMIP6 models in the representation of surface NO_3 (Figure 7) and surface CHL (Figure 20). Annual cycle comparison shows a misrepresentation of spring and fall blooms due to an overall negative bias in NO_3 and CHL. As such, five out of six CMIP6 models underestimate spring bloom CHL and fall bloom CHL increases are entirely absent in all CMIP6 models (Figure 5). With phytoplankton blooms in the northern North Atlantic accounting for up to 500 g C $year^{-1}$ (Mueter et a., 2009), an accurate representation of the magnitude of such events is of immense importance in climate research. Further, misrepresentations of surface CHL and nutrients in the euphotic zone in productive regions can lead to biased carbon export calculations and thereby affect the accuracy of future climate projections.

Underestimation of NO_3 within the euphotic zone is partially caused by underestimation of NO_3 at depth (Figure 6) which inhibits enough nutrients being mixed upwards into the euphotic zone. Despite this, depth profiles in the Northern North Atlantic (Figure 6, top panels) show that surface CHL is underestimated, but most model-Argo Δs lie within the natural variability of the observations. Euphotic Zone NO_3, however, is strongly underestimated and values are far outside the standard deviation which is assumed here to represent the natural variability. This discrepancy between large NO_3 biases and smaller CHL biases, is interesting as one would expect CHL bias patterns to follow those of NO_3 in eutrophic environments where CHL concentrations within the euphotic zone are strongly dependent on nutrient supply from below. A similar example is visible in the ARTC annual cycle, for which CHL is overestimated by the MPI-ESM1.2, but NO_3 is constantly underestimated (Figure 5, top two panels from left). This pattern raises the question how CHL concentrations can increase without enough NO_3 being available and indicates that the relationship between NO_3 and CHL might be biased in the model parameterization or that CHL concentrations have been tuned to fit observations at the surface. While tuning is a common procedure in ocean modelling, it can lead to seemingly correct results but for the wrong reasons. For example, reduced CHL biases at the surface do not imply similarly

good agreement at depth. Models GFDL-ESM4, MPI-ESM1.2, and NorESM5 include tuning for BGC properties (Séférian et al., 2020, supplement, Stock et al., 2020; Paulsen et al. 2017 and Mauritsen et al., 2019; Tjiputra et al., 2020).

4.1.2 Biases in ocean physics

Surface bias maps were used to visualize TEMP and SAL biases located in the Greenland and Norwegian Sea, together with a biased representation of Gulf Stream physics. Depth profiles show that CMIP6 models misrepresent physical properties in the mixed layer (top 150 m) and at intermediate depths (500-1200 m).

In the northern North Atlantic, these biases indicate a possible misrepresentation of the location and strength of both the NAC and the Iceland Current by all CMIP6 models. This can lead to biases in modelled deep water formation and can cause a misplacement of Labrador Sea Water (LSW) which is important for the transport of oxygen to the entire Atlantic (Koelling et al., 2022). In ARCT, the preformed component of ΔNO_3 depth profiles is much larger than the remaining part due to BGC (Figure 11, top panels). This indicates that a bias in the model physics persists and that it explains a large fraction of the NO_3 bias in this region.

The southern oligotrophic North Atlantic, physical biases were more pronounced at intermediate depth. Strong TEMP and SAL overestimations were found, suggesting issues with intermediate water heat transport and misrepresentations of intermediate water masses. Similarly, as in ARCT, NO_3 biases are more strongly affected by model physics than by BGC processes, as indicated by large underestimations in preformed NO_3 (Figure 11, horizontal middle and bottom panels).

MOI modelled physics compare significantly better to the observations than CMIP6 does: Spatial representation for TEMP and SAL at the surface, including a much more accurate Gulf Stream separation, is superior to those of CMIP6 (Figure 8, Figure 19). Additionally, model-Argo-Δs are much lower than those of CMIP6 for most provinces and properties (Figure 6, Figure 17). Biases between WOA and MOI physics are expected to be minimal, as MOI-GLO12 assimilates WOA13 data.

4.1.3 Misplacement of the Gulf Stream

The province surrounding the Gulf Stream, GFST, marks the transition zone between the eutrophic northern North Atlantic and the oligotrophic southern North Atlantic. WOA data indicate that surface NO_3 concentrations increase starting at 45°N towards the North. This NO_3 transition zone is either absent or misplaced by CMIP6 models (Figure 7). MOI places the transition zone at 40°N and slightly overestimates the increase in surface NO_3 with latitude.

CHL satellite data places the transition zone for CHL further south, at around 32°N (Figure 20). 3/6 CMIP6 models capture the location well, but magnitudes and patterns of surface CHL still differ significantly from the observations.

GFST average Δ-depth profile patterns are more similar to those of the southern North Atlantic, indicating that primary productivity is not the main driver for biases in this province (Figure 17). In fact, the largest biases are located at 500-1200 m depth, as seen for southern provinces (Figure 6, bottom panels).

Modelled physical properties in the Gulf Stream region showed large biases for surface TEMP. This indicates a general misplacement of the Gulf Stream, as well as misrepresented Gulf Stream properties. Further, the width of the Gulf Stream at the surface is biased, as seen in Figure 8, by an overestimation on the western side of the Gulf Stream turning into an underestimation towards the east.

Numerical representation of the Gulf Stream separation is notoriously difficult. While higher resolution (10 km at the equator) enhances the modeling of the Gulf Stream position, achieving the accurate pathway remains elusive. Despite 70 years of literature on the Gulf Stream separation, a definitive understanding of the dynamics governing the current's trajectory has yet to be attained. Because separation dynamics remain unclear, developing precise modeling techniques for the Gulf Stream separation at increased resolutions remains a challenge (Bryan et al., 2007; Schoonover et al., 2017; Schoonover et al., 2016).

4.1.4 Does a good agreement of CHL in the mixed layer imply a similar agreement throughout the entire euphotic zone?

CMIP6 modelled mixed layer (0-6 m) CHL lies outside of the natural variability of the observations for some models in northern provinces and for most models in southern provinces (Figure 6). However, IPSL-CM6A modelled mixed layer CHL agrees well with the observations in WTRA and shows strong over- and underestimations at deeper depths in the euphotic zone (0-110 m). Overall, agreement between modelled CHL and observations improves throughout the euphotic zone in eutrophic provinces, but large biases persist within the euphotic zone in oligotrophic regions (Figure 6).

This means that even if modelled surface concentrations appear to fit the observations well, biases can be found below the mixed layer.

4.1.5 Is euphotic zone CHL affected by the representation of NO_3 below the euphotic zone?

Comparatively to the oligotrophic southern half of the study domain, North of 45°N surface nitrate is elevated predominantly due to stronger vertical mixing, through which nutrients are transported to the euphotic zone (Schulz et al., 2022). CHL concentrations are expected to increase with an increased availability of nitrate (given light and temperature are favourable for photosynthesis to occur) in these regions (Sigman & Hain, 2012). Because of this linkage between euphotic zone CHL and deep water NO_3, model performance for these properties is expected to be similar.

However, annual cycles in eutrophic northern provinces showed mismatches during spring blooms between underestimated NO_3 and overestimated CHL for some models (MPI-ESM1.2, CanESM5) (Figure 5).

Further, largest NO_3 biases occur at approximately at 100 m depth and below 500 m (Figure 6). These biases are model underestimations that exceed the estimated natural variability of Argo observations in most provinces. CHL in contrast is found to be underestimated but remain within the natural variability (Figure 6). CMIP6 model NorESM2 even overestimates NO_3 at 0-500m while still underestimating euphotic zone CHL in northern North Atlantic provinces (Figure 6, Figure 17).

Thereby, euphotic zone CHL appears to be independent of both NO_3 within the euphotic zone and NO_3 concentrations below 110 m for most CMIP6 models in eutrophic provinces. CHL is often diagnosed in models by phytoplankton biomass and thereby a function of nutrient concentration, light availability, temperature, and others. While biomass is a common proxy for CHL, it is important to note that the cellular content of phytoplankton and the ratio with respect to carbon are variable. For this reason, CHL a biased estimator for biomass (Huot et al., 2007).

4.1.6 Is the representation of ocean properties within the euphotic zone the same as it is below?

For many properties and in multiple provinces, largest biases are located at intermediate depth and thereby below the euphotic zone. Nevertheless, surface and euphotic zone concentrations were shown to vary significantly between models and observations. No evidence was found that a low misfit of modelled properties within the euphotic zone guarantees similarly good agreement below 110 m or vice versa.

4.2 Investigate the interrelation between biases in BGC and physical models.

4.2.1 Does a misfit in the model physics affect agreement of BGC variables?

Biases in model physics were found to indicate a misrepresentations of general circulation patterns and important physical features in the northern North Atlantic. Such biases can likely affect modelled BGC, as the physical model components of Earth System Models are coupled to the BGC model component (e.g. Dunne et al., 2020 for GFDL-ESM4), but further, because BGC process parameterizations are based on physical properties (e.g. photosynthesis being TEMP-dependent).

Throughout all analyses shown in this paper, intermediate depth (500-1200 m) and surface waters have been identified as the regions for largest biases amongst models, provinces, and properties.

NO_3 biases at intermediate depth are anticorrelated with OXY biases (Figure 22), indicating a linkage between the variables. Strong NO_3-OXY anticorrelations could be an

indication for a misrepresentation of respiration, however this is not further investigated here.

Additionally, strong anticorrelations were found for biases of TEMP and NO_3 in oligotrophic provinces and SARC. It was further shown that the preformed part of ΔNO_3 depth profiles explains a larger fraction of the overall bias than the BGC driven part in both ARCT and NATR (Figure 11), which indicates a significant effect of biased model physics on BGC model performance.

4.2.2 Does a more complex BGC model or increased spatial resolution guarantee better model performance?

Biased physics and circulation patterns can be the result of coarse model resolution which affects ocean mixing (Griffies et al., 2009). While this thesis only focusses on the effect of spatial/horizontal resolution on model performance, due to the method of vertical interpolation, vertical resolution also significantly contributes to model performance. As such, vertical resolution determines the model's ability to capture vertical property gradients and the representation of vertical processes, including upwelling, mixing, and stratification.

MOI models resolve ocean BGC with 1/4° resolution while ocean physics even have a resolution of 1/12°. CMIP6 models included in this study resolve 100 km at the equator (~1°), except for GFDL-ESM4 which has the highest CMIP6 resolution of 1/2°. As MOI models overall strongly outcompete CMIP6 models and the GFDL-ESM4 ranks best amongst CMIP6, model resolution is likely explaining some biases found here. However, data-assimilation is likely a more significant factor contributing to MOI models' performance, rather than the high resolution.

These findings could be interpreted as controverse to Laurent et al (2021) who report that increased model resolution does not guarantee better model performance, however, amongst the five CMIP6 models with spatial resolutions of 1°, performance varies significantly with property and location. For this reason, we conclude that resolution alone does not improve model performance.

A possible explanation for why resolution alone does not improve model performance, is that extensive tuning becomes increasingly difficult in a high-resolution model due to an extensive computational cost. For example, HAMOCC (MPI-ESM1.2, NorESM2) uses tuning of biological model properties (e.g. plankton decay and grow rates) to accomplish reasonable biomass estimates for N_2-fixers and their fixation rates. This method requires additional sensitivity experiments to evaluate how the uncertainties in these parameters affect model results (Paulsen et al., 2017). With increased resolution these methods become more computationally expensive. Therefore, a lower resolution model might be thoroughly tuned and perform better than the higher resolution version (Moreno-Chamarro et al., 2022).

Another contributing factor to model performance is BGC model complexity. CMIP6 models NorESM2 and CanESM5 are the only two CMIP6 models that use fixed Redfield Ratios for all variables and are set up to only resolve one phytoplankton type. Variations in Redfield Ratios can be implemented by adjusting the stoichiometric ratios of carbon, nitrogen, and phosphorus in phytoplankton and other components of the marine ecosystem, based on environmental conditions.

NorESM2 and CanESM5 are also the two models that have the overall highest misfits. Further, CanESM5 only includes nitrogen as nutrients, while the other CMIP6 models resolve up to five (GFDL-ESM4) different nutrients (Séférian et al., 2020, supplement).

An average taken over all model ranks classified GFDL-ESM4 as the best CMIP6 performing model in the North Atlantic (Figure 12). This model does not just resolve at twice the horizontal resolution of most other CMIP6 models, it also uses one of the most complex BGC models (COBALTv2).

COBALTv2 resolves five nutrients, four phytoplankton types, three zooplankton types and it is the only BGC model that explicitly models bacteria, while CanESM5 only includes nitrate and one phyto-/zooplankton type each. However, the most complex BGC model in terms of number of nutrients and phytoplankton types, is CMCC-ESM2 (5 nutrients, 4 phytoplankton, 4 zooplankton, bacteria).

In conclusion, there are indications that, provided the ocean physics are accurately represented possibly through increased spatial resolution, a more complex BGC model should perform better than a basic BGC model.

4.3 Assessment of a CMIP6 model ensemble performance with respect to individual models including data-assimilation models.

For different analyses shown in this thesis different biases were discussed. Some biases were found to persist amongst all six CMIP6 models (e.g. TEMP underestimation within euphotic zone and OXY overestimation from 0-2000m in NADR Figure 5, Figure 6), but also individually different biases have been reported (e.g. surface NO_3 concentrations and OXY concentrations at 0-2000 m in WTRA Figure 7, Figure 6). Depending on the location and property, model performance varies and not one CMIP6 model is always better than others (Figure 12). However, MOI models tend to outcompete CMIP6 models in most instances. An overall average rank identifies MPI-ESM1.2 as the model with the lowest misfit for data comparison in the euphotic zone, as found by Laurent et al., 2021. GFDL-ESM4 can be titled the “best performing CMIP6 model” below the euphotic zone and for the entire water column which agrees with a study by Rickard et al. (2022) who labelled GFDL-ESM4 as one of the best 3 ESM out of 32 models (16 CMIP5 and 16 CMIP6 models) in an assessment of physical and BGC components for the Ross Sea.

An ensemble mean of all six CMIP6 models used here shows that there is large variability amongst CMIP6 models. Depending on province and property, the ensemble mean either outperforms CMIP6 or falls behind individual models. Therefore, a CMIP6 ensemble mean does not necessarily provide the best result. Further, the ensemble mean falls behind MOI model performance (Figure 12).

Amongst all models discussed here, CanESM5 can be identified as the model with the overall highest misfit both in the upper water column (0-110 m) and below (Figure 12). CanESM5 shows strong physical biases in the surface and below (strong TEMP bias in Labrador Sea and south of Greenland Figure 8; overestimated SAL and TEMP at depth in WTRA Figure 6), together with a poor representation of BGC variables (e.g. surface NO_3 and BGC properties at 0-2000 m in NADR Figure 7, Figure 6). While such biased model

physics are likely affecting the representation of BGC properties in CanESM5, some bias might be the result of a comparatively simpler BGC model, which lacks accurate representation of marine BGC.

4.4. Uncertainties

Analyses shown in this paper come with uncertainties which need to be assessed and put in relation to the identified model biases.

First, different time periods are used for comparisons of observational data sets from WOA and (BGC-)Argo float data. As expected and regardless of the different times, the comparison of WOA data to Argo data is always better than any CMIP6 model comparison with Argo. Moreover, WOA data mostly falls within one STD of Argo observations which indicates that the estimated natural variability of Argo data is large enough to enclose WOA climatologies. It further illustrates the similarity of Argo and WOA which means that the Argo program is dense enough to be used for this model evaluation.

Furthermore, Lavoie et al. (2013) found surface NO_3 and CHL trends from 1970-2000 to be negligible which means that the large time period of 1900-2021 used for WOA BGC properties, might not be affecting observational data comparisons. WOA and Argo AOU concentrations do differ. 65-70% of this difference is explained by OXY measurements and around 30-40% by TEMP (Figure 23). Similarly, MOI and Argo AOU do not compare well (Figure 5), but differences are larger for TEMP than they are for OXY (Figure 6). While Argo OXY sensor errors due to a slow response time are a known issue (Bittig et al., 2014), OXY sensors are the most accurate BGC sensors with an Argo fleet-wide accuracy of 1% or better (Johnson et al. 2017a, Claustre et al., 2020). Nevertheless, compared to CTD sensors, OXY response times are significantly longer (Bennett & Huaide, 1986). However, challenges in accuracy arising from these extended response times are predominantly encountered in gradients (Gordon et al., 2020). This means, that here reported MOI AOU biases are partially caused by TEMP biases which could possibly be the result of averaging (discussed below).

Second, CMIP6 data was averaged over a 30-year time span (1990-2009) roughly covering that of availability of Argo observations. This was done to allow models to capture

interannual variability on time scales <30 years. Therefore, some bias between CMIP6 and Argo is to be expected. +/- one standard deviation of Argo floats was used to estimate natural variability. As discussed in 2.5 Standard Deviation (STD), in this thesis, *natural* variability denotes any variability, including anthropogenic signals and variability on varying timescales. The 30-year averaged CMIP6 data should fall within this range for good agreement between models and observations. Similarly, MOI model data were averaged over 12 years (2009-2021) for consistency. MOI models ought to proficiently replicate natural variability by incorporating data assimilation techniques.

Third, (BGC-)Argo coverage is still sparse in some regions, especially for NO_3 and OXY. While only provinces with relatively good coverage were included, large gaps in the province-wide coverages still persist. If (BGC-)Argo profile coverage were less clustered in each province, the average model performance might look different.

5 CONCLUSION

Six CMIP6 models together with data-assimilating MOI models were evaluated against (BGC-)Argo and WOA observations. Comparisons identified multiple biases amongst CMIP6 models which are summarized and quantified in Table 3. While it was demonstrated that MOI models overall performed better than CMIP6, MOI modelled AOU has larger misfits than any CMIP6 model in SARC (Figure 12) which are in part due to differences in TEMP between MOI and observations (possibly caused by averaging).

For CMIP6 models, modelled NO_3 was found to be inaccurate at 0-2000 m depth throughout the study domain, while other properties' biases were more localized at the surface and at intermediate depths. Additionally, physical property comparison in the northern North Atlantic showed a misrepresentation of circulation patterns including the NAC and the Gulf Stream, as well as possible issues with deep water formation in the Greenland and Norwegian Sea. It was further shown that CMIP6 bias patterns do not always agree (surface NO_3 and OXY depth profiles in WTRA) and that properties can be linked (TEMP vs. NO_3 biases) or be unrelated (CHL and NO_3 biases).

These findings were discussed to be similar to other model evaluations (see 4.1. Evaluating whether global ocean models can reproduce an accurate state of North Atlantic properties.). The resemblances observed across various studies suggest that the methods employed in this context are adequate. Additionally, they indicate that the uncertainties discussed are relatively smaller in comparison to the reported findings.

As expected, this evaluation demonstrates that a data-assimilating model with a high spatial resolution outperforms current global ocean models used in CMIP6. It further shows that CMIP6 models are unable to accurately represent North Atlantic properties, which demonstrates the urgent need for model improvement to facilitate accurate model applications.

Additionally, it was shown that the (BGC-)Argo fleet is a useful tool to locate and discuss model biases. BGC-Argo data can be used to understand and explain BGC processes in more detail which could improve BGC model parameterization (Wang & Fennel, 2023) and reduce biases found here. However, enhancements in spatial Argo coverage could mitigate uncertainties stemming from scant data and the biases introduced by clustered

profile locations. Improved spatial coverage also strengthens the reliability of the data and refines the accuracy of assessments.

To augment the precision of my analyses, I prioritize advancing my analytical methodologies. As such, I would choose an alternative method for separating my study domain into provinces. Employing machine learning techniques such as clustering approaches holds promise in this regard, facilitating the automatic identification of distinct regions based on data patterns rather than predefined boundaries, such as the here used Longhurst provinces. This would not only streamline the process but also potentially reveal novel insights by uncovering previously overlooked spatial nuances.

Table 3: Model-Argo bias overview.

	CMIP6-Argo Biases	**MOI-Argo Biases**	**WOA-Argo Discrepancies**
Northern North Atlantic Spring and Fall Bloom	• -0.7 mg/m^3 (CanESM5) • No second CHL increase in fall • Timing of bloom biased	• -0.3 mg/m^3 • Underestimated fall bloom CHL increase	
Northern North Atlantic NO_3	• -10 µmol/kg (CanESM5) to 8 µmol/kg (NorESM2)	• Overestimations within Euphotic Zone (<110 m) up to 5 µmol/kg • Below 200 m ±1 µmol/kg	• ±1 µmol/kg
TEMP within Euphotic Zone in NADR	• -1 °C (GFDL-ESM2) to -5 °C (CanESM5)	• ~ ±1 °C	• < ±1 °C

	CMIP6-Argo Biases	MOI-Argo Biases	WOA-Argo Discrepancies
Euphotic Zone OXY	• Overestimations >80 µmol/kg (CanESM5, IPSL-CM6A)	• Overestimations: 5-30 µmol/kg	• Overestimations: 5-20 µmol/kg
Intermediate Depth (500-1500 m) OXY	• Large variability amongst model biases: -50 to 85 µmol/kg	• Overestimations: 10-30 µmol/kg	• -2 to 10 µmol/kg
Intermediate Depth (500-1500 m) Physics	• SAL: -0.3 (CanESM5) to 1.1 (NorESM2, CMCC-ESM2) • TEMP overestimations reaching up to 4 °C (NorESM2, CMCC-ESM2)	• SAL < | 0.1 | • TEMP overestimations reaching up to 2 °C in GFST	• SAL < | 0.1 | • TEMP: -0.5 to 0.1 °C

APPENDICES

A.1 Longhurst province processing

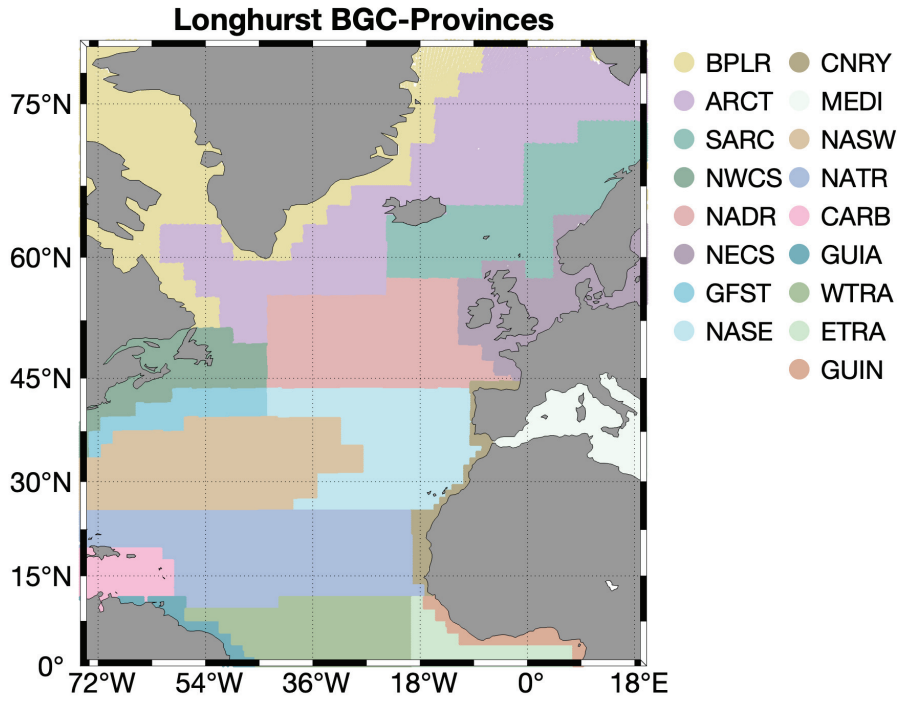


Figure 13: Original Longhurst provinces in the North Atlantic.

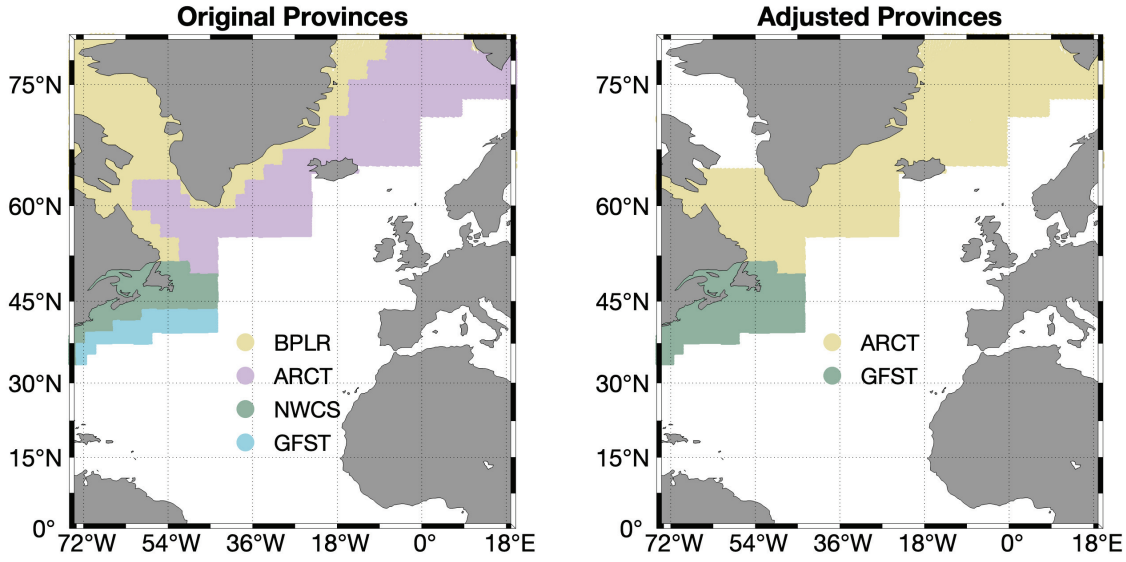


Figure 14: Left panel shows provinces that were adjusted/combined due to (BGC-)Argo coverage and shelves. Right panel shows adjusted provinces used in the evaluation.

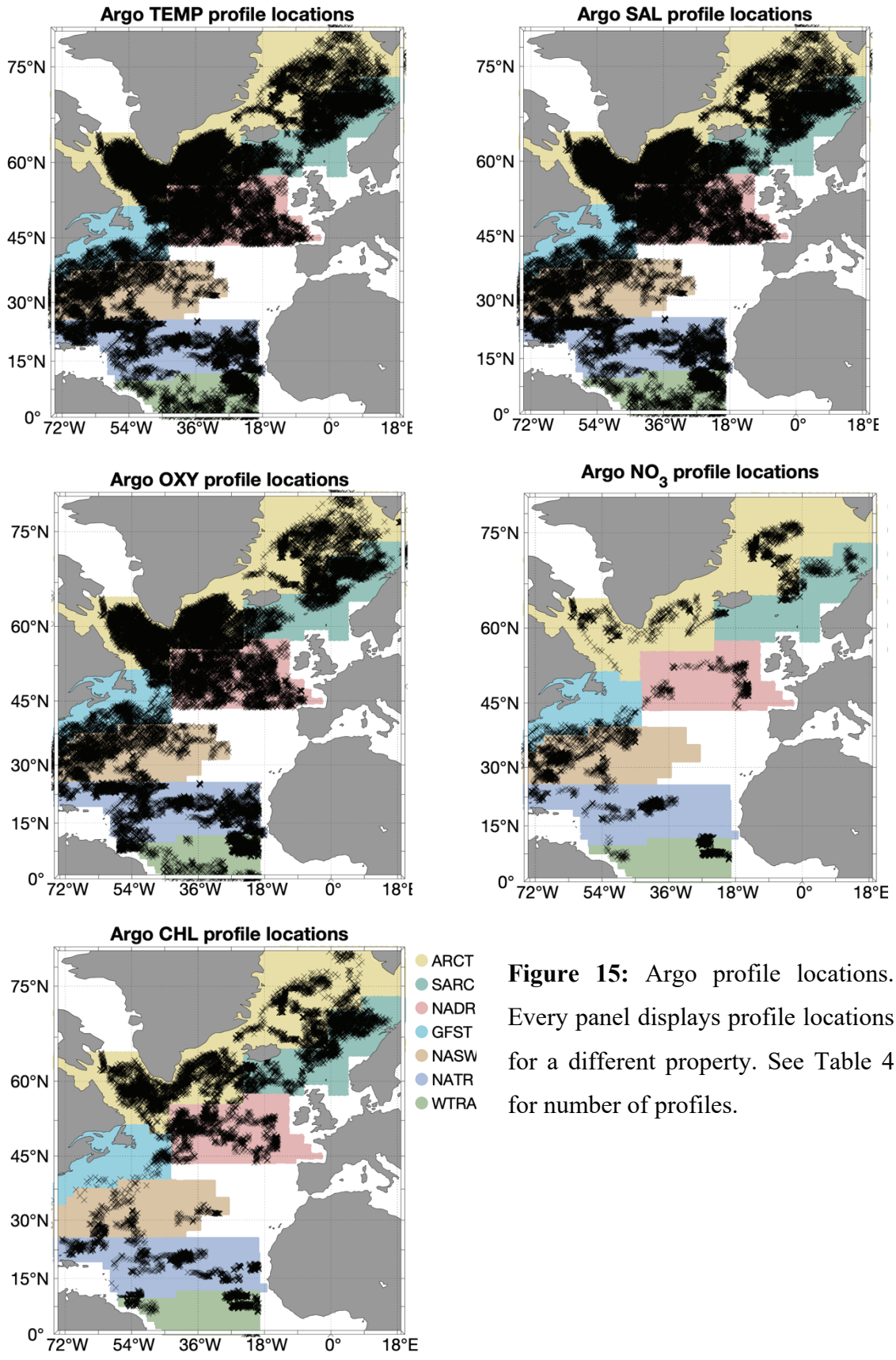


Figure 15: Argo profile locations. Every panel displays profile locations for a different property. See Table 4 for number of profiles.

Table 4: Number of profiles used in this evaluation per property and province. Corresponding profile locations in Figure 15.

	CHL	NO_3	OXY	SAL	TEMP
ARCT	5193	1458	12075	16521	16524
SARC	2413	355	2901	4527	4527
NADR	1907	381	4178	6346	6640
GFST	183	241	1178	1459	1459
NASW	416	1154	2445	2859	2939
NATR	2353	976	4850	5831	5831
WTRA	1219	1103	2546	2927	2927

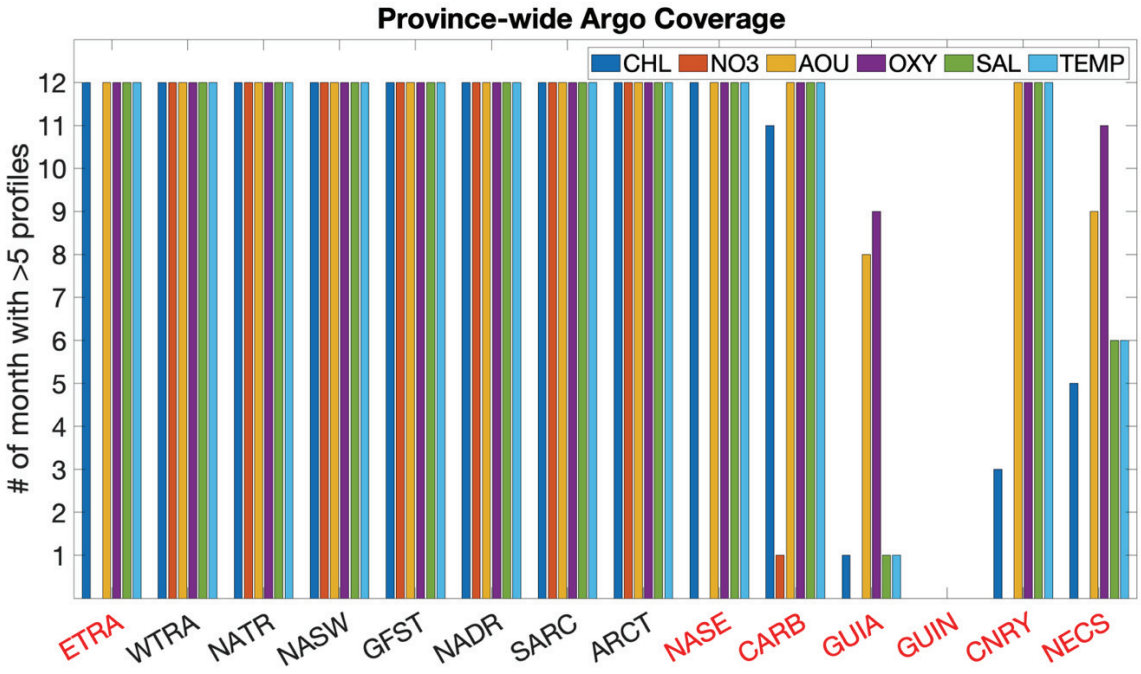
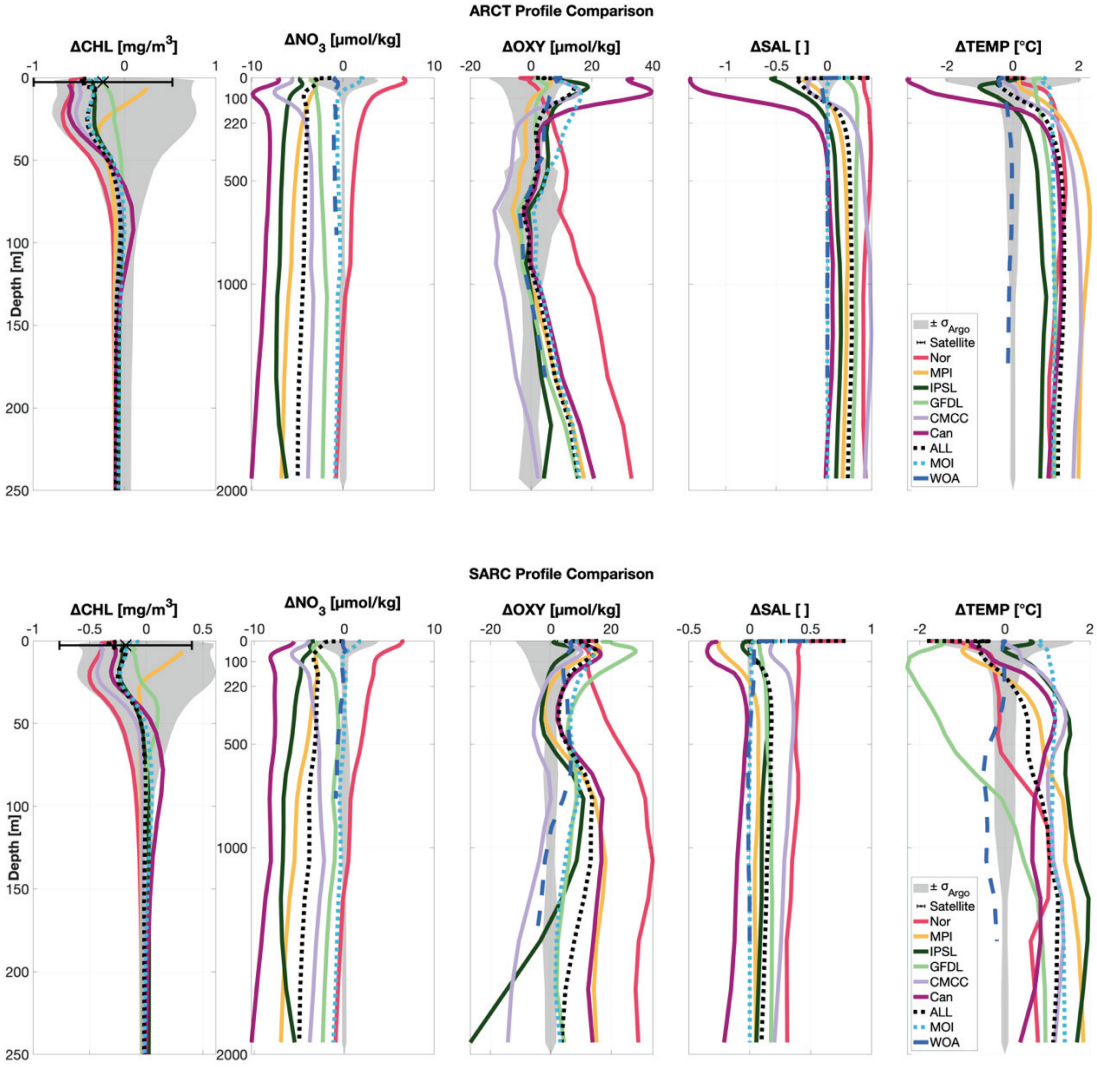


Figure 16: Histogram of number of month with more than five good quality-flagged (1/2/5/8) Argo profiles per property (different colours, see legend) for each province. Provinces that exceed the threshold are WTRA, NATR, NASW, GFST, NADR, SARC & ARCT. Provinces marked in red are left out of the evaluation.

A.2 Δ-Profile Comparisons



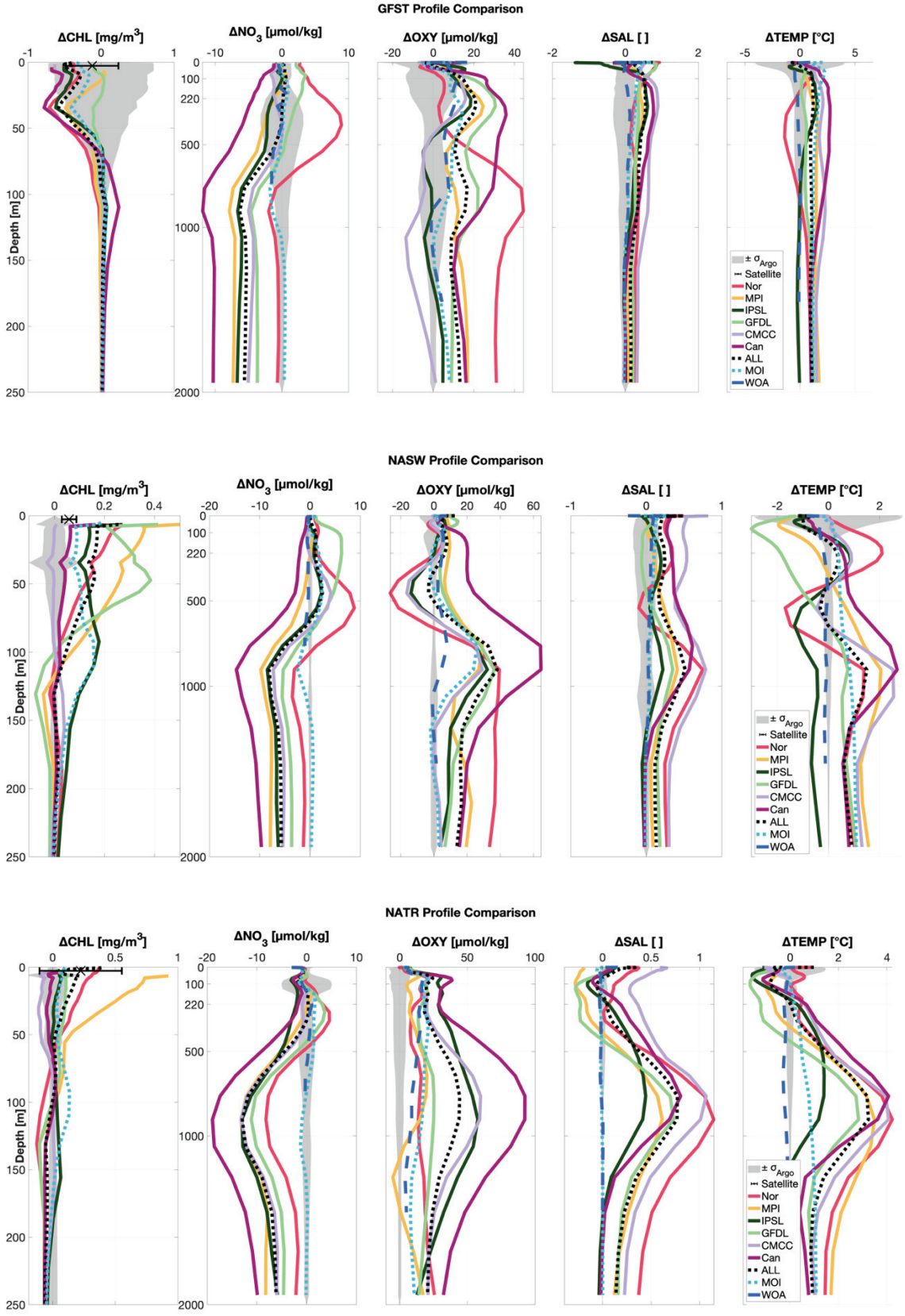

GFST Profile Comparison
ΔCHL [mg/m3]
ΔNO3 [μmol/kg]
ΔOXY [μmol/kg]
ΔSAL [ ]
ΔTEMP [°C]
Depth [m]
± σArgo
Satellite
Nor
MPI
IPSL
GFDL
CMCC
Can
ALL
MOI
WOA
NASW Profile Comparison
ΔCHL [mg/m3]
ΔNO3 [μmol/kg]
ΔOXY [μmol/kg]
ΔSAL [ ]
ΔTEMP [°C]
Depth [m]
± σArgo
Satellite
Nor
MPI
IPSL
GFDL
CMCC
Can
ALL
MOI
WOA
NATR Profile Comparison
ΔCHL [mg/m3]
ΔNO3 [μmol/kg]
ΔOXY [μmol/kg]
ΔSAL [ ]
ΔTEMP [°C]
Depth [m]
± σArgo
Satellite
Nor
MPI
IPSL
GFDL
CMCC
Can
ALL
MOI
WOA

Figure 17: For all 5 figures above. Average delta (model-Argo) depth profiles for CHL, NO_3, OXY, SAL & TEMP, each figure for a different province. Black x with ±1 standard deviation indicates satellite CHL, Argo data in dashed black, WOA in dashed blue, MOI in light blue dotted line, CMIP6 models in remaining solid colours, and the black dotted line is an ensemble of all CMIP6 models (ALL). WOA, MOI, and CMIP6 data from locations of Argo profiles only. Grey shading displays ±1 standard deviation of Argo data. CHL profiles shown to 250m depth, other properties to 2000m.

A.3 Annually Averaged Surface Biases and Surface Concentrations

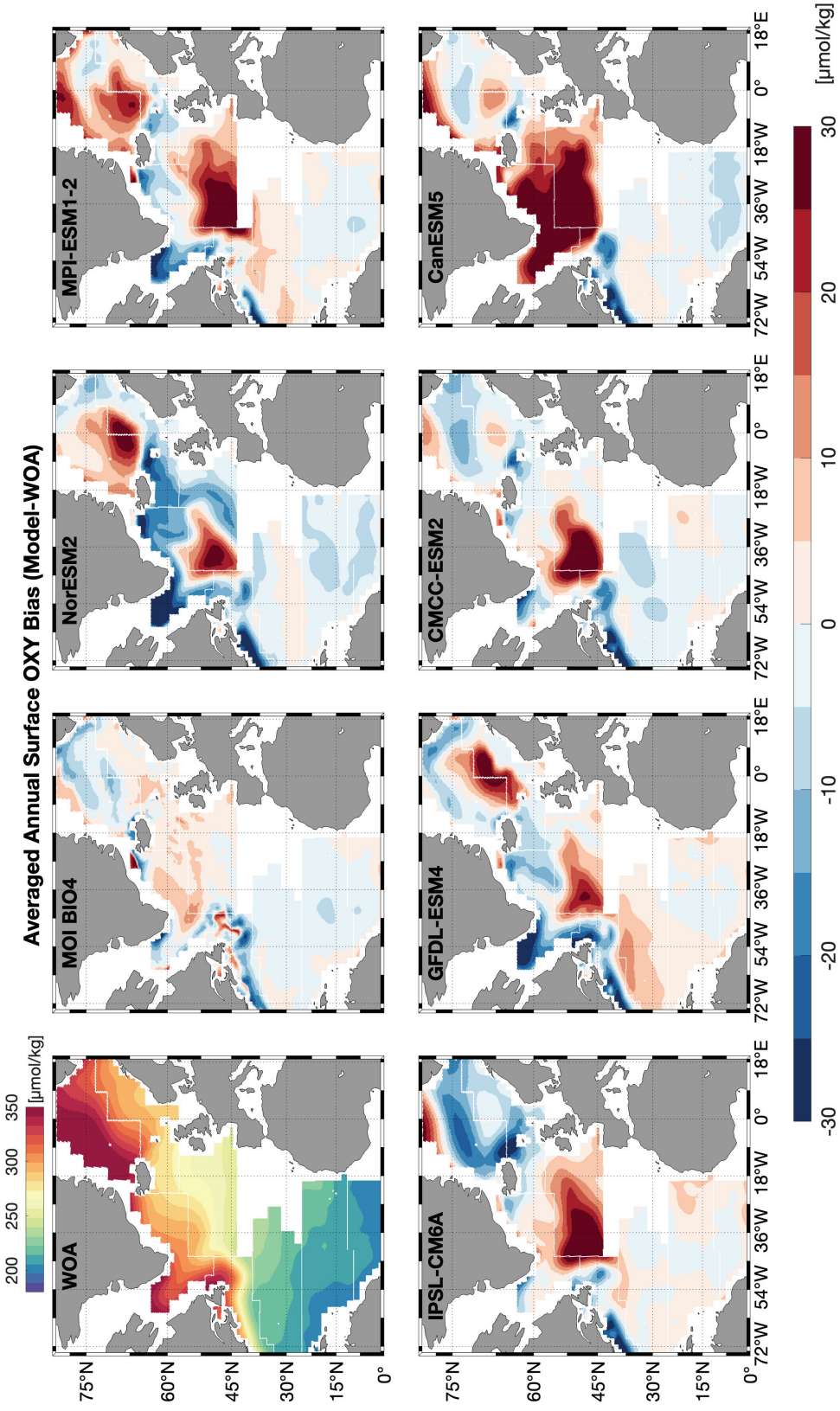


Figure 18: Annually averaged surface OXY [µmol/kg] bias (Model-WOA). Top row left panel shows WOA annual average from 1900-2017. Top row second from the left is data-assimilating model MOI from 2009-2021. Remaining panels are CMIP6 models, averaged over 1990-2019. Surface refers to an average over 0-6m.

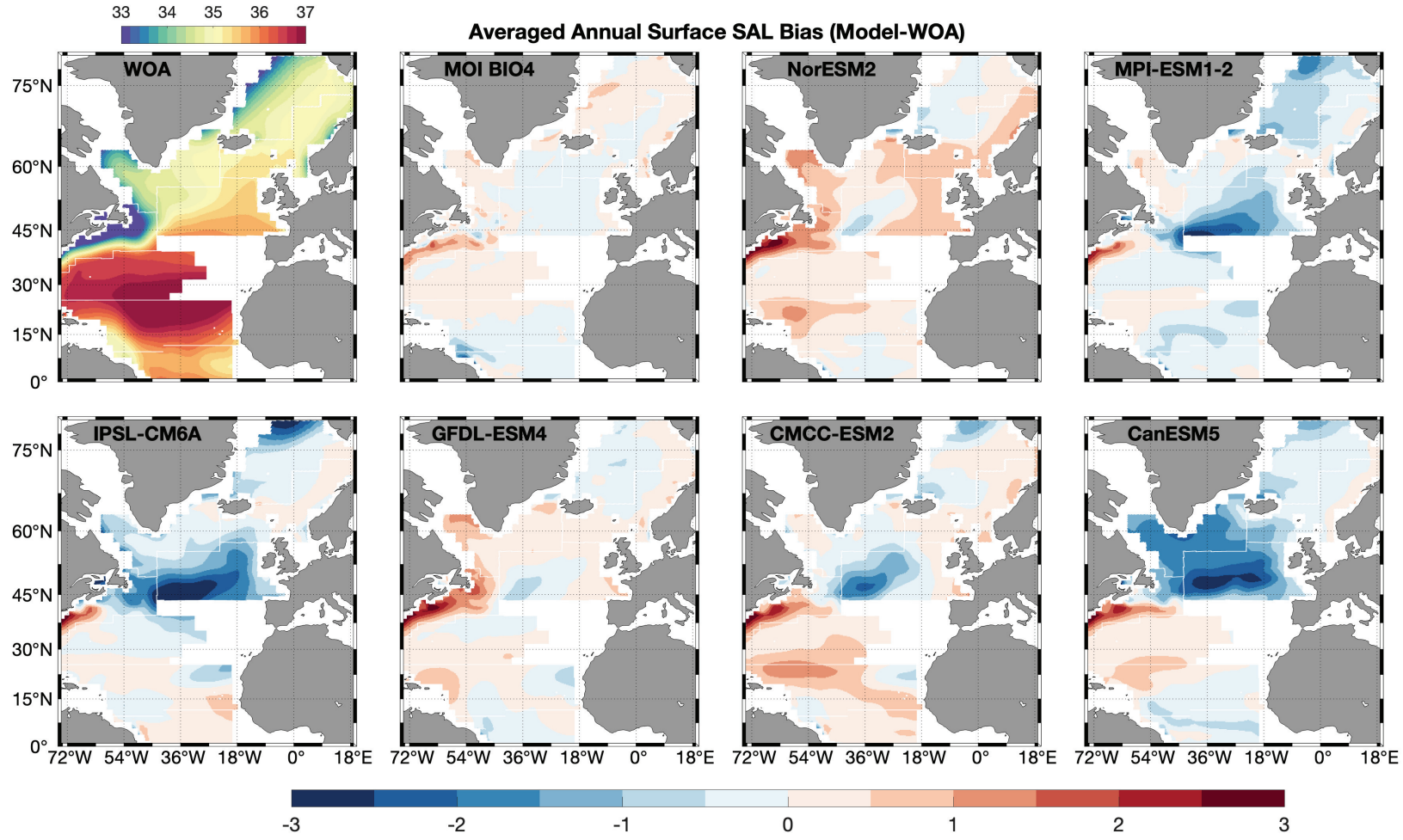


Figure 19: Annually averaged surface SAL bias (Model-WOA). Top row left panel shows WOA annual average from 1900-2017. Top row second from the left is data-assimilating model MOI from 2009-2021. Remaining panels are CMIP6 models, averaged over 1990-2019. Surface refers to an average over 0-6m.

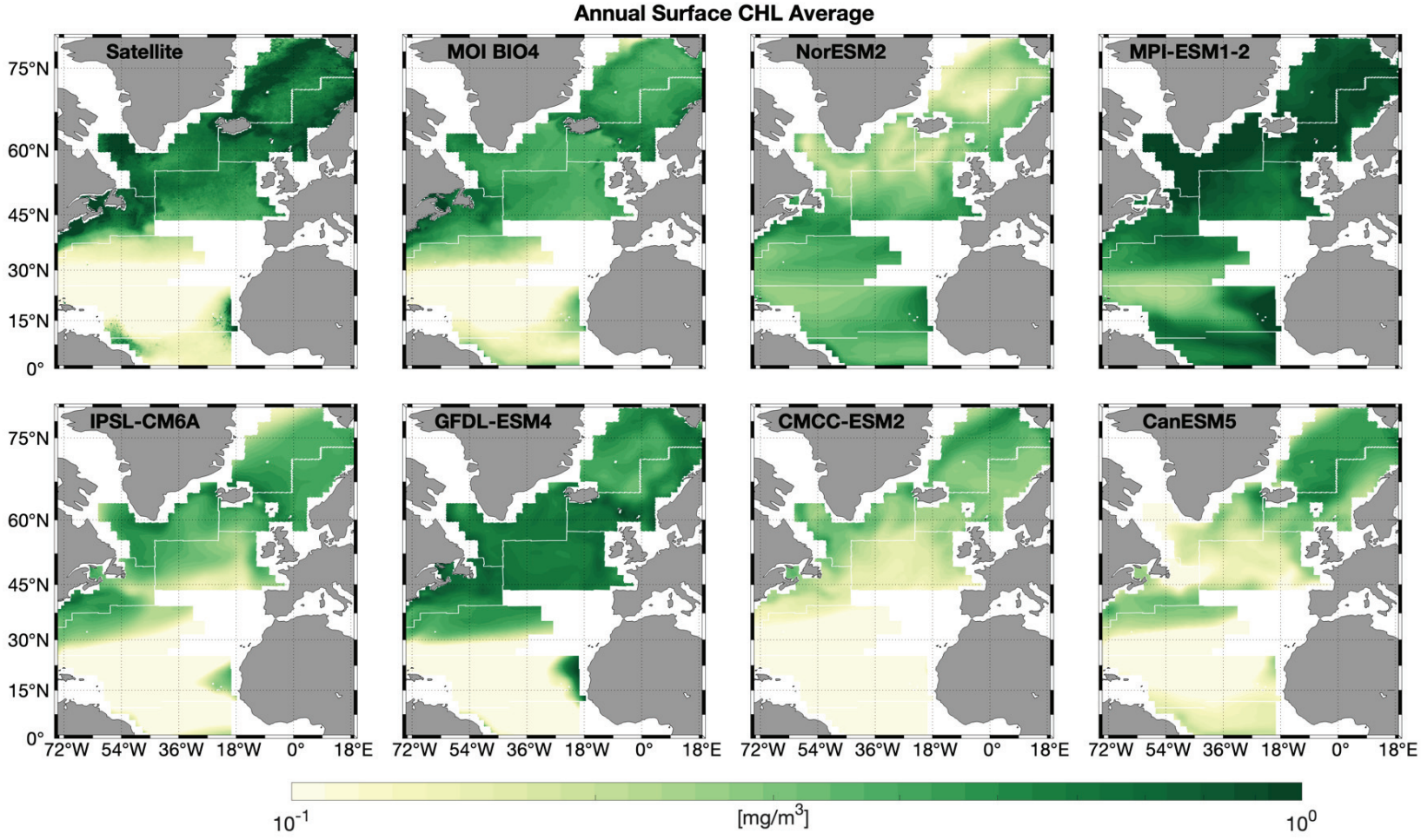


Figure 20: Annual surface CHL [mg/m^3] average. Top row left panel shows satellite annual average from 2003-2021. Top row second from the left is data-assimilating model MOI from 2009-2021. Remaining panels are CMIP6 models, averaged over 1990-2019. Surface model data refers to an average over 0-6m.

A.4 Bias Correlation

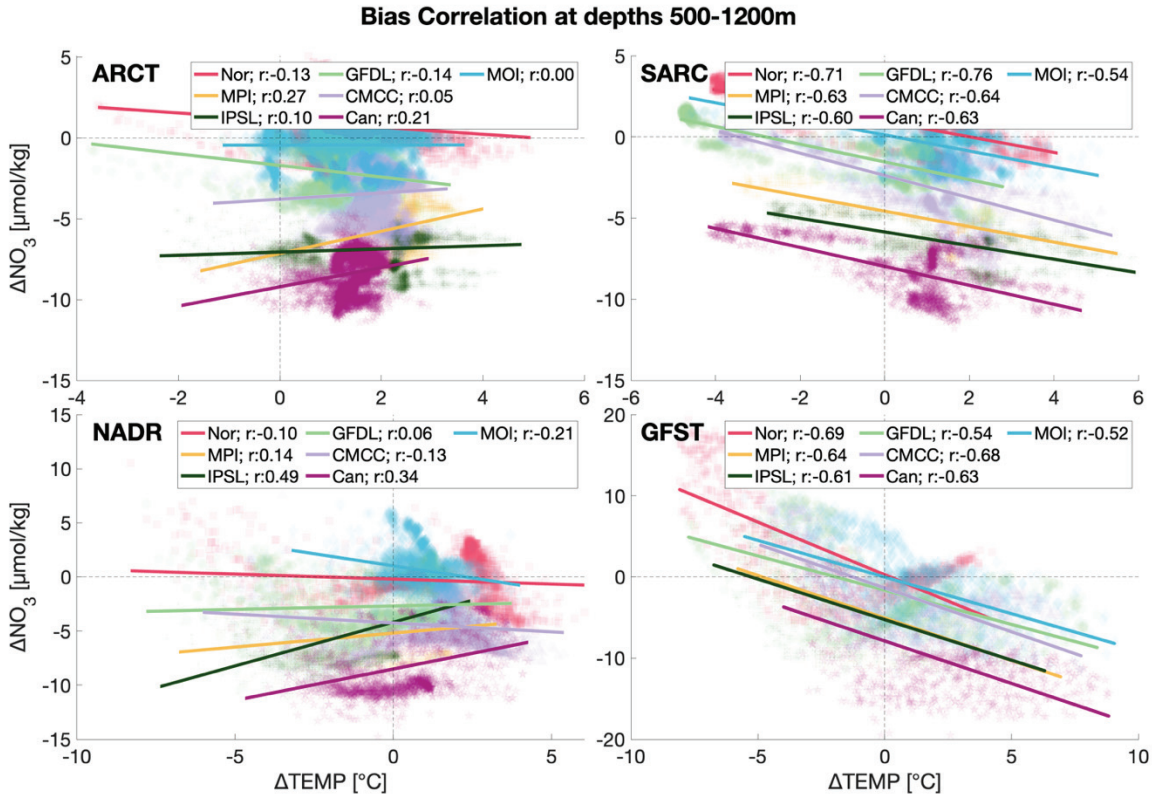


Figure 21: TEMP biases vs. NO_3 biases at 500-1200m for provinces ARCT, SARC, NADR & GFST. Coloured lines indicate regression for each model, Pearson's correlation coefficients r are given in legend. Negative r indicates anticorrelation.

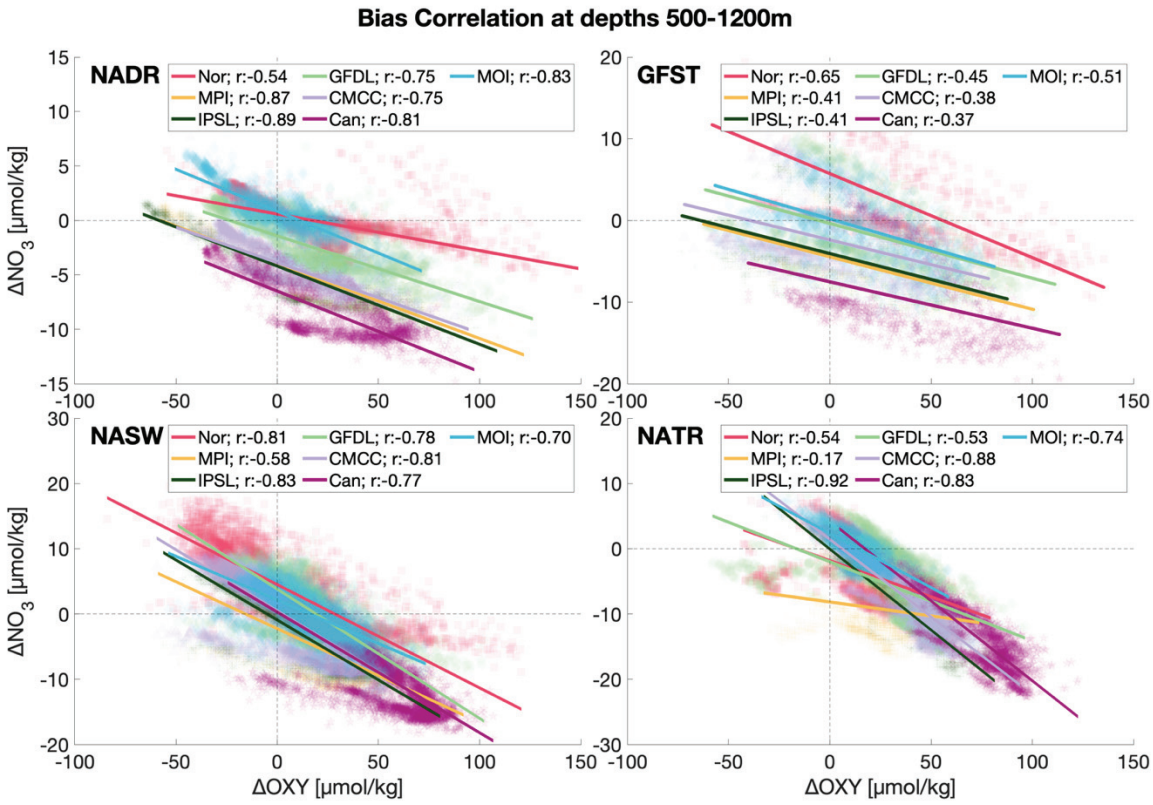


Figure 22: OXY biases vs. NO_3 biases at 500-1200m for provinces NADR, GFST, NASW & NATR. Coloured lines indicate regression for each model, Pearson's correlation coefficients r are given in legend. Negative r indicates anticorrelation.

6.5 AOU Mismatch Test

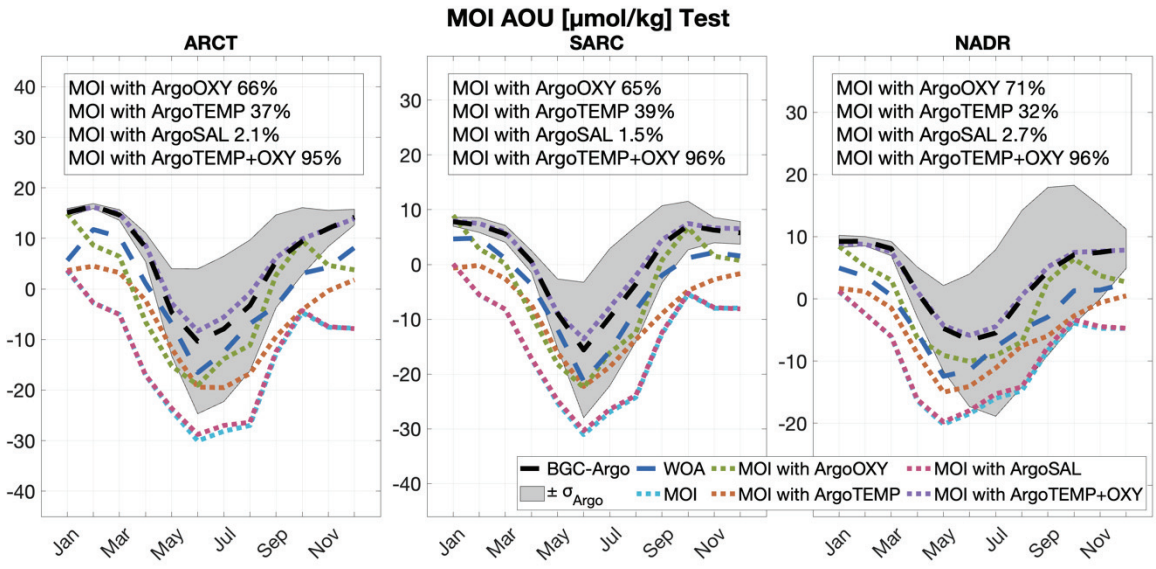


Figure 23: Test explaining mismatch between MOI and (BGC-)Argo AOU in provinces ARCT, SARC & NADR (panels from left to right). AOU was calculated from MOI OXY, TEMP & SAL and then compared to calculations using (BGC-)Argo data for either OXY, TEMP or SAL or both TEMP and OXY. Percentages listed indicate how much of the mismatch is being explained by each property (e.g. MOI AOU calculations using Argo OXY explain 65-71% of the mismatch between Argo and MOI AOU). Percentages were calculated form averages over 0-110m depth and were further averaged over time which leads to some rounding errors.

REFERENCES

Bakun, A. (1990). Global climate change and intensification of coastal ocean upwelling. Science, 247(4939), 198-201.

Bennett, A. S. and Huaide, T.: CTD time-constant correction, Deep-Sea Res., 33, 1425–1438, 1986.

Bennett, A. F., 2005: Inverse modeling of the ocean and atmosphere. Cambridge University Press.

Bittig HC, Fiedler B, Scholz R, Krahmann G, Körtzinger A. 2014. Time response of oxygen optodes on profiling platforms and its dependence on flow speed and temperature. Limnol. Oceanogr. Methods 12:617–36.

Bonan, G. B. and Doney, S. C.: Climate, ecosystems, and planetary futures: The challenge to predict life in Earth system models, Sci- ence, 359, eaam8328, https://doi.org/10.1126/science.aam8328, 2018.

Boyd, P. W., & Trull, T. W. (2007). Understanding the export of biogenic particles in oceanic waters: Is there consensus?. Progress in Oceanography, 72(4), 276-312.

Brandt, P., Alory, G., Awo, F. M., Dengler, M., Djakouré, S., Imbol Koungue, R. A., ... & Rouault, M. (2023). Physical processes and biological productivity in the upwelling regions of the tropical Atlantic. Ocean Science, 19(3), 581-601.

Bryan, F. O., M. W. Hecht, and R. D. Smith (2007), Resolution convergence and sensitivity studies with North Atlantic circula- tion models. Part I: The western boundary current system. Ocean Modelling, 16, 141–159.

Broecker, W. S., & Peng, T. H. (1982). Tracers in the Sea (Vol. 690). Palisades, New York: Lamont-Doherty Geological Observatory, Columbia University.

Chassignet, E., & Marshall, D. (2008). Gulf Stream separation in numerical ocean models. Geophysical Monograph Series, 177, 39–61.

Claustre, H., Johnson, K. S., & Takeshita, Y. (2020). Observing the global ocean with biogeochemical-Argo. Annual review of marine science, 12, 23-48.

Dengg, J. (1993), The problem of Gulf Stream separation: A baro- tropic approach. J. Phys. Oceanogr., 23, 2182–2200.

Dickson, B., Yashayaev, I., Meincke, J., Turrell, B., Dye, S., & Holfort, J. (2002). Rapid freshening of the deep North Atlantic Ocean over the past four decades. Nature, 416(6883), 832-837.

Ducklow, H. and McCallister, S.: The biogeochemistry of carbon dioxide in the coastal oceans, in: The Sea, Vol 13: The global coastal ocean. Multiscale interdisciplinary processes, edited by Robinson, A. R. and Brink, K. H., pp. 269–315, John Wiley & Sons, New York, 2004.

Dunne, J. P., Horowitz, L. W., Adcroft, A. J., Ginoux, P., Held, I. M., John, J. G., et al. (2020). The GFDL Earth System Model Version 4.1 (GFDL-ESM 4.1): Overall coupled model description and simulation characteristics. Journal of Advances in Modeling Earth Systems, 12, e2019MS002015. https://doi.org/10.1029/2019MS002015

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., & Taylor, K. E. (2016). Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. Geoscientific Model Development, 9(5), 1937-1958.

Fu, W., Moore, J. K., Primeau, F., Collier, N., Ogunro, O. O., Hoffman, F. M., & Randerson, J. T. (2022). Evaluation of ocean biogeochemistry and carbon cycling in CMIP earth system models with the International Ocean Model Benchmarking (IOMB) software system. Journal of Geophysical Research: Oceans, 127, e2022JC018965. https://doi.org/10.1029/2022JC018965

Frenzel, Hartmut (co-first author), Sharp, Jonathan D. (co-first author), Fassbender, Andrea J., Plant, Josh N., Maurer, Tanya L., Takeshita, Yui, Nicholson, David P., & Gray, Alison R. (2021). BGC-Argo-Mat: A MATLAB toolbox for accessing and visualizing Biogeochemical Argo data (v1.0.1). Zenodo. https://doi.org/10.5281/zenodo.5015054

Garcia H.E., T.P. Boyer, O.K. Baranova, R.A. Locarnini, A.V. Mishonov, A. Grodsky, C.R. Paver, K.W. Weathers, I.V. Smolyar, J.R. Reagan, D. Seidov, M.M. Zweng (2019). World Ocean Atlas 2018: Product Documentation. A. Mishonov, Technical Editor.

Ghil, M., and P. Malanotte-Rizzoli, 1991: Data assimilation in meteorology and oceanography. Advances in

Geophysics, 33, 141–266.

Gordon, C., Fennel, K., Richards, C., Shay, L. K., and Brewster, J. K.: Can ocean community production and respiration be determined by measuring high-frequency oxygen profiles from autonomous floats?, Biogeosciences, 17, 4119–4134, https://doi.org/10.5194/bg-17-4119-2020, 2020.

Griffies, S. M., Adcroft, A. J., Banks, H., Böning, C. W., Chassignet, E. P., Danabasoglu, G., ... & White, L. (2009). Problems and prospects in large-scale ocean circulation models. Proceedings of OceanObs, 9, 410-431.

Gruber, N., Clement, D., Carter, B. R., Feely, R. A., Van Heuven, S., Hoppema, M., ... & Wanninkhof, R. (2019). The oceanic sink for anthropogenic CO2 from 1994 to 2007. Science, 363(6432), 1193-1199.

Holt, J., Hyder, P., Ashworth, M., Harle, J., Hewitt, H. T., 60 Liu, H., New, A. L., Pickles, S., Porter, A., Popova, E., Allen, J. I., Siddorn, J., and Wood, R.: Prospects for im- proving the representation of coastal and shelf seas in global ocean models, Geoscientific Model Development, 10, 499–523, https://doi.org/10.5194/gmd-10-499-2017, 2017.

Honjo S.. Ittekkot V., Shhäfer P., Honjo S., Depetris P. J., et al. Fluxes of particles to the interior of the open oceans, Particle Flux in the Ocean., 1996 ChichesterWiley and Sons, 91-145

Hoteit, I., Luo, X., Bocquet, M., Kohl, A., & Ait-El-Fquih, B. (2018). Data assimilation in oceanography: Current status and new directions. New frontiers in operational oceanography, 465-512.

Huot, Y., Babin, M., Bruyant, F., Grob, C., Twardowski, M. S., & Claustre, H. (2007). Does chlorophyll a provide the best index of phytoplankton biomass for primary productivity studies?. Biogeosciences discussions, 4(2), 707-745.

Ilyina, T., Six, K. D., Segschneider, J., Maier-Reimer, E., Li, H., & Núñez-Riboni, I. (2013). Global ocean biogeochemistry model HAMOCC: Model architecture and performance as component of the MPI-Earth system model in different CMIP5 experimental realizations. Journal of Advances in Modeling Earth Systems, 5(2), 287-315.

Jayne, S. R., Roemmich, D., Zilberman, N., Riser, S. C., Johnson, K. S., Johnson, G. C., & Piotrowicz, S. R. (2017). The Argo program: present and future. Oceanography, 30(2), 18-28.

Jickells, T. D., An, Z. S., Andersen, K. K., Baker, A. R., Bergametti, G., Brooks, N., ... & Torres, R. (2005). Global iron connections between desert dust, ocean biogeochemistry, and climate. science, 308(5718), 67-71.

Johnson KS, Plant JN, Coletti LJ, Jannasch HW, Sakamoto CM, et al. 2017a. Biogeochemical sensor performance in the SOCCOM profiling float array. J. Geophys. Res. Oceans 122:6416–36

Kiehl, J. T., and P. R. Gent (2004), The community climate system model, version two. J. Clim., 17, 3666–3682.

Koelling, J., Atamanchuk, D., Karstensen, J., Handmann, P., & Wallace, D. W. (2022). Oxygen export to the deep ocean following Labrador Sea Water formation. Biogeosciences, 19(2), 437-454.

Laurent, A., Fennel, K., & Kuhn, A. (2021). An observation-based evaluation and ranking of historical Earth system model simulations in the northwest North Atlantic Ocean. Biogeosciences, 18(5), 1803-1822.

Lavoie, D., Lambert, N., den Mustapha, S., and van der Baaren, A.: Projections of Future Physical and Biogeochemical Conditions in the Northwest Atlantic from CMIP5 Global Climate Models, Tech. rep., Vol. 285, 2013. a, b, c, d

Laws, E. A., Falkowski, P. G., Smith Jr, W. O., Ducklow, H., & McCarthy, J. J. (2000). Temperature effects on export production in the open ocean. Global biogeochemical cycles, 14(4), 1231-1246.

Lellouche, J.-M., Greiner, E., Bourdallé-Badie, R. B., Garric, G., Melet, A., Drévillon, M. ... & Le Traon, P.-Y. (2021). The Copernicus global 1/12 oceanic and sea ice GLORYS12 reanalysis. Frontiers in Earth Science, 9, 698876.

Longhurst, A., Sathyendranath, S., Platt, T., & Caverhill, C. (1995). An estimate of global primary production in the ocean from satellite radiometer data. Journal of plankton Research, 17(6), 1245-1271.

Masson-Delmotte, V., Zhai, P., Pirani, A., Connors, S. L., Péan, C., Berger, S., ... & Zhou, B. (2021). Climate change 2021: the physical science basis. Contribution of working group I to the sixth assessment report of the intergovernmental panel on climate change, 2.

Mauritsen, T., Bader, J., Becker, T., Behrens, J., Bittner, M., Brokopf, R., ... & Roeckner, E. (2019). Developments in the MPI-M Earth System Model version 1.2 (MPI-ESM1. 2) and its response to increasing CO2. Journal of Advances in Modeling Earth Systems, 11(4), 998-1038.

Meehl, G. A., Karl, T., Easterling, D. R., Changnon, S., Pielke Jr, R., Changnon, D., ... & Zwiers, F. (2000). An introduction to trends in extreme weather and climate events: observations, socioeconomic impacts, terrestrial ecological impacts, and model projections. Bulletin of the American Meteorological Society, 81(3), 413-416.

Meehl, G. A., Covey, C., Delworth, T., Latif, M., McAvaney, B., Mitchell, J. F., ... & Taylor, K. E. (2007). The WCRP CMIP3 multimodel dataset: A new era in climate change research. Bulletin of the American meteorological society, 88(9), 1383-1394.

Miller, C. B., & Wheeler, P. A. (2012). Biological oceanography. John Wiley & Sons.

Moreno-Chamarro, E., Caron, L. P., Loosveldt Tomas, S., Vegas-Regidor, J., Gutjahr, O., Moine, M. P., ... & Vidale, P. L. (2022). Impact of increased resolution on long-standing biases in HighResMIP-PRIMAVERA climate models. Geoscientific Model Development, 15(1), 269-289.

Mueter, F. J., Broms, C., Drinkwater, K. F., Friedland, K. D., Hare, J. A., Hunt Jr, G. L., ... & Taylor, M. (2009). Ecosystem responses to recent oceanographic variability in high-latitude Northern Hemisphere ecosystems. Progress in Oceanography, 81(1-4), 93-110.

Muller-Karger, F. E.: The importance of continental margins in the global carbon cycle, Geophysical Research Letters, 32, L01 602, https://doi.org/10.1029/2004GL021346, 2005.

Munk, W. H. (1950). On the wind-driven ocean circulation. Journal of Atmospheric Sciences, 7(2), 80-93.

O'Neill, B. C., Tebaldi, C., Van Vuuren, D. P., Eyring, V., Friedlingstein, P., Hurtt, G., ... & Sanderson, B. M. (2016). The scenario model intercomparison project (ScenarioMIP) for CMIP6. Geoscientific Model Development, 9(9), 3461-3482.

Parekh, P., Follows, M. J., Dutkiewicz, S., & Ito, T. (2006). Physical and biological regulation of the soft tissue carbon pump. Paleoceanography, 21(3).

Paulsen, H., Ilyina, T., Six, K. D., & Stemmler, I. (2017). Incorporating a prognostic representation of marine nitrogen fixers into the global ocean biogeochemical model HAMOCC. Journal of Advances in Modeling Earth Systems, 9(1), 438-464.

Perner, K., Moros, M., Jansen, E., Kuijpers, A., Troelstra, S. R., & Prins, M. A. (2018). Subarctic Front migration at the Reykjanes Ridge during the mid-to late Holocene: evidence from planktic foraminifera. Boreas, 47(1), 175-188.

Pickart, R. S., Torres, D. J., & Clarke, R. A. (2002). Hydrography of the Labrador Sea during active convection. Journal of Physical Oceanography, 32(2), 428-457.

Pommier, J., Gosselin, M., & Michel, C. (2009). Size-fractionated phytoplankton production and biomass during the decline of the northwest Atlantic spring bloom. Journal of plankton research, 31(4), 429-446.

Puerta, P., Johnson, C., Carreiro-Silva, M., Henry, L. A., Kenchington, E., Morato, T., ... & Orejas, C. (2020). Influence of water masses on the biodiversity and biogeography of deep-sea benthic ecosystems in the North Atlantic. Frontiers in Marine Science, 239.

Reygondeau, G., Longhurst, A., Martinez, E., Beaugrand, G., Antoine, D., & Maury, O. (2013). Dynamic biogeochemical provinces in the global ocean. Global Biogeochemical Cycles, 27(4), 1046-1058.

Riahi, K., Van Vuuren, D. P., Kriegler, E., Edmonds, J., O'neill, B. C., Fujimori, S., ... & Tavoni, M. (2017). The Shared Socioeconomic Pathways and their energy, land use, and greenhouse gas emissions implications: An overview. Global environmental change, 42, 153-168.

Richter, J. H., Anstey, J. A., Butchart, N., Kawatani, Y., Meehl, G. A., Osprey, S., & Simpson, I. R. (2020). Progress in simulating the quasi-biennial oscillation in CMIP models. Journal Geophysical Research: Atmospheres, 125, e2019JD032362. https://doi.org/10.1029/2019JD032362

Roemmich, D., Johnson, G. C., Riser, S., Davis, R., Gilson, J., Owens, W. B., ... & Ignaszewski, M. (2009). The Argo Program: Observing the global ocean with profiling floats. Oceanography, 22(2), 34-43.

Roemmich, D., Alford, M. H., Claustre, H., Johnson, K., King, B., Moum, J., ... & Yasuda, I. (2019). On the future of Argo: A global, full-depth, multi-disciplinary array. Frontiers in Marine Science, 6, 439.

Sabine, C. L., Feely, R. A., Gruber, N., Key, R. M., Lee, K., Bullister, J. L., ... & Rios, A. F. (2004). The oceanic sink for anthropogenic CO2. science, 305(5682), 367-371.

Sanders, R., Henson, S. A., Koski, M., Christina, L. D. L. R., Painter, S. C., Poulton, A. J., ... & Martin, A. P. (2014). The biological carbon pump in the North Atlantic. Progress in Oceanography, 129, 200-218.

Schmitz Jr, W. J., & McCartney, M. S. (1993). On the north Atlantic circulation. Reviews of Geophysics, 31(1), 29-49.

Schoonover, J., Dewar, W. K., Wienders, N., & Deremble, B. (2017). Local sensitivities of the Gulf Stream separation. Journal of Physical Oceanography, 47(2), 353-373.

Schoonover, J., Dewar, W., Wienders, N., Gula, J., McWilliams, J. C., Molemaker, M. J., ... & Yeager, S. (2016). North Atlantic barotropic vorticity balances in numerical models. Journal of Physical Oceanography, 46(1), 289-303.

Schulz, K., Lincoln, B., Povazhnyy, V., Rippeth, T., Lenn, Y.-D., Janout, M., et al. (2022). Increasing nutrient fluxes and mixing regime changes in the eastern Arctic Ocean. Geophysical Research Letters, 49, e2021GL096152. https://doi.org/10.1029/2021GL096152

Séférian, R., Berthet, S., Yool, A., Palmieri, J., Bopp, L., Tagliabue, A., ... & Yamamoto, A. (2020). Tracking improvement in simulated marine biogeochemistry between CMIP5 and CMIP6. Current Climate Change Reports, 6(3), 95-119.

Siegel DA, Buesseler KO, Doney SC, Sailley SF, Behrenfeld MJ, Boyd PW. Global assessment of ocean carbon export by combining satellite observations and food-web models. Glob Biogeochem Cycles. 2014;28:181–96.

Sigman, D. M., & Boyle, E. A. (2000). Glacial/interglacial variations in atmospheric carbon dioxide. Nature, 407(6806), 859-869.

Sigman, D. M. & Hain, M. P. (2012) The Biological Productivity of the Ocean. Nature Education Knowledge 3(10):21

Stock, C. A., Dunne, J. P., Fan, S., Ginoux, P., John, J., Krasting, J. P., ... & Zadeh, N. (2020). Ocean biogeochemistry in GFDL's Earth System Model 4.1 and its response to increasing atmospheric CO2. Journal of Advances in Modeling Earth Systems, 12(10), e2019MS002043.

Stoer, A. C., Takeshita, Y., Maurer, T. L., Demeaux, C. B., Bittig, H. C., Boss, E., ... & Fennel, K. (2023). A census of quality-controlled Biogeochemical-Argo float measurements. Frontiers in Marine Science.

Stommel, H. (1948). The westward intensification of wind-driven ocean currents. Eos, Transactions American Geophysical Union, 29(2), 202-206.

Sverdrup, H. U., & Munk, W. H. (1947). Wind, sea and swell: Theory of relations for forecasting (No. 601). Hydrographic Office.

Taylor, K. E., Stouffer, R. J., and Meehl, G. A.: An Overview of Cmip5 and the Experiment Design, B. Am. Meteorol. Soc., 93, 485–498, 2012.

Tjiputra, J. F., Schwinger, J., Bentsen, M., Morée, A. L., Gao, S., Bethke, I., ... & Schulz, M. (2020). Ocean biogeochemistry in the Norwegian Earth system model version 2 (NorESM2). Geoscientific Model Development, 13(5), 2393-2431.

Vantrepotte, V., & Mélin, F. (2009). Temporal variability of 10-year global SeaWiFS time-series of phytoplankton chlorophyll a concentration. ICES Journal of Marine Science, 66(7), 1547-1556.

Volk, T., & Hoffert, M. I. (1985). Ocean carbon pumps: Analysis of relative strengths and efficiencies in ocean-driven atmospheric CO2 changes. The carbon cycle and atmospheric CO2: Natural variations Archean to present, 32, 99-110.

Wang, B., & Fennel, K. (2023). An assessment of vertical carbon flux parameterizations using backscatter data from BGC Argo. Geophysical Research Letters, 50(3), e2022GL101220.

Weiss, R. F. (1970, August). The solubility of nitrogen, oxygen and argon in water and seawater. In Deep sea research and oceanographic abstracts (Vol. 17, No. 4, pp. 721-735). Elsevier.

Wunsch, C., 1996: The ocean circulation inverse problem. Cambridge University Press, Cambridge, UK.

https://web.archive.org/web/20060218120705/http://www.giss.nasa.gov/research/briefs/legrande_01/

https://argo.ucsd.edu/how-do-floats-work/

Data:

https://esgf-node.llnl.gov/search/cmip6/

ftp://ftp.ifremer.fr/ifremer/argo

https://www.ncei.noaa.gov/access/world-ocean-atlas-2018/

https://www.mercator-ocean.eu/en/solutions-expertise/accessing-digital-data/

https://oceandata.sci.gsfc.nasa.gov/directdataaccess/Level-3%20Mapped/Aqua-MODIS