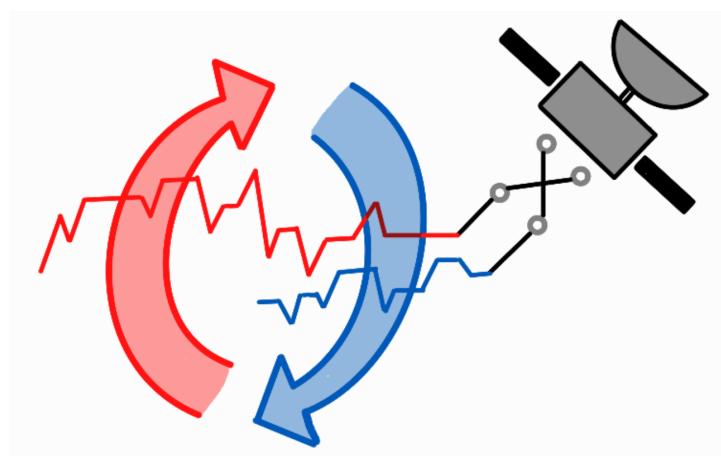


# Using Machine Learning and Satellite-Observable Variables to Predict Abyssal Ocean Circulation

Tom Cowperthwaite, Aline Van Driessche, Sharan Maiya, Nina Baranduin, Emilio Luz-Ricca



Department of Earth Sciences  
University of Cambridge  
United Kingdom  
20/03/2024

## Abstract

The meridional overturning circulation (MOC) plays a fundamental role in climate regulation through its transport of heat, nutrients, oxygen, and carbon. At the margins of Antarctica, winds and sea ice formation lead to cold, saline, and therefore dense water to sink to the ocean floor, forming Antarctic Bottom Water (AABW). Increasing temperatures due to anthropogenic climate change may limit the formation of these abyssal waters, potentially affecting a key part of global ocean circulation. Unfortunately, tracking this phenomenon through traditional methods is logically challenging, due to the various difficulties of studying the Southern Ocean. In this report, we demonstrate that machine learning techniques can predict abyssal MOC strength using only satellite-observable variables instead. We train a suite of models for this task using the “Estimating the Circulation and Climate of the Ocean” (ECCO) state estimate, obtaining state-of-the-art performance. We incorporate the “Australian Community Climate and Earth System Simulator Ocean Model” (ACCESS), a high-resolution numerical ocean circulation model; and observational “Rapid Climate Change-Meridional Overturning Circulation and Heatflux Array” (RAPID) data, a cross-basin sensor array that directly measures the Atlantic MOC strength. Our experiments indicate an approximately linear relationship between satellite-observable variables and abyssal MOC strength. We additionally demonstrate the utility of observational data for predicting long-range oceanic dependencies through the integration of RAPID, and show that a deep learning model is able to accurately capture latitude-invariant features for MOC strength prediction. Through these experiments, we present a methodology for predicting abyssal circulation, which will be instrumental in informing climate policy and enabling further oceanographic research.

## Report Structure

In the **introduction**, we contextualize our project, introduce the research questions that we aim to address, and summarize our novel contributions. In the **methods**, we introduce datasets, define our prediction goal, explain preprocessing and training details, and elaborate on how our models will be evaluated. In the **results**, we establish a baseline methodology, replicate and extend the current state-of-the-art, and explore three novel experiments. In the **conclusion**, we summarize our progress towards our initial goals, briefly discuss limitations and future work, and discuss the broader implications of our findings.

Each (sub-)section is the product of dedicated efforts by a specific lead author. This author is acknowledged at the start of each respective (sub-)section in *italics*. The lead author was responsible for not only the writing of their (sub-)section, but also the design and implementation of experiments, as well as the presentation of results. This ensures each part of the report benefits from the focused and unique perspective of its primary contributor.

## List of Abbreviations

- **1D-CNN**: 1-Dimensional Convolutional Neural Network
- **2D-CNN**: 2-Dimensional Convolutional Neural Network
- **AABW**: Antarctic Bottom Water
- **ACC**: Antarctic Circumpolar Current
- **ACCESS**: Australian Community Climate and Earth System Simulator
- **AMOC**: Atlantic Meridional Overturning Circulation
- **CNN**: Convolutional Neural Network
- **COSIMA**: Consortium for Ocean-Sea Ice Modelling in Australia
- **ECCO**: Estimating the Circulation and Climate of the Ocean
- **GPU**: Graphics Processing Unit
- **GRU**: Gated Recurrent Unit
- **IPCC**: Intergovernmental Panel on Climate Change
- **LLC90**: lat-lon-cap 90 grid
- **LSTM**: Long Short-Term Memory
- **MAE**: Mean Absolute Error
- **MAPE**: Mean Absolute Percentage Error
- **MITgcm**: Massachusetts Institute of Technology General Circulation Model
- **ML**: Machine Learning
- **MLP**: Multi-Layer Perceptron
- **MOC**: Meridional Overturning Circulation
- **MOM5**: Modular Ocean Model
- **MSE**: Mean Squared Error
- **NASA**: National Aeronautics and Space Administration
- **NN**: Neural Network
- **NOC**: National Oceanography Centre
- **OBP**: Ocean Bottom Pressure
- **PO.DAAC**: Physical Oceanography Distributed Active Archive Center
- **RAPID-MOCHA**: Rapid Climate Change-Meridional Overturning Circulation and Heatflux Array
- **ReLU**: Rectified Linear Unit
- **RNN**: Recurrent Neural Network
- **RMSE**: Root Mean Square Error
- **SSH**: Sea Surface Height
- **SSS**: Sea Surface Salinity
- **SST**: Sea Surface Temperature
- **Sv (unit)**: Sverdrups ( $1 \text{ Sv} \equiv 10^6 \text{ m}^3\text{s}^{-1}$ )
- **ZWS**: Zonal Wind Stress

# Contents

|          |  |           |
|----------|--|-----------|
| <b>1</b> | <b>Introduction</b>  | <b>4</b>  |
| <b>2</b> | <b>Methods</b>   | <b>5</b>  |
| 2.1      | Numerical Models . . . . .   | 5         |
| 2.2      | Ocean Basins and Latitudes of Interest . . . . .                       | 5         |
| 2.3      | Satellite-observable Variables . . . . .                               | 5         |
| 2.4      | MOC Strength Calculation . . . . .                                     | 6         |
| 2.5      | Variable Preprocessing . . . . .                                       | 9         |
| 2.6      | Machine Learning Training Details . . . . .                            | 9         |
| 2.7      | Evaluation Metrics . . . . .   | 9         |
| <b>3</b> | <b>Results &amp; Discussion</b>  | <b>9</b>  |
| 3.1      | Baseline Modelling: Linear Regression . . . . .                        | 10        |
| 3.2      | Replication of Solodoch <i>et al.</i> . . . . .                        | 10        |
| 3.2.1    | Training Details . . . . .   | 11        |
| 3.2.2    | Model Performance . . . . .  | 11        |
| 3.2.3    | Limitation: Loss of Important Signal . . . . .                         | 12        |
| 3.3      | Model Advancements . . . . .   | 12        |
| 3.3.1    | Trend and seasonality . . . . .  | 13        |
| 3.3.2    | 1D-CNNs . . . . .  | 14        |
| 3.3.3    | 2D-CNNs . . . . .  | 15        |
| 3.4      | Experiment 1: Incorporating Observational Data (RAPID) . . . . .       | 16        |
| 3.5      | Experiment 2: Modelling All Latitudes in the Southern Ocean . . . . .  | 17        |
| 3.5.1    | Linear Regression . . . . .  | 18        |
| 3.5.2    | Learning Latitude-Invariant Features . . . . .                         | 19        |
| 3.6      | Experiment 3: Using High-resolution Simulation Data (ACCESS) . . . . . | 20        |
| <b>4</b> | <b>Conclusion</b>  | <b>20</b> |
| <b>5</b> | <b>Code and Data Availability</b>                                      | <b>21</b> |
| <b>A</b> | <b>Appendices</b>  | <b>27</b> |
| A.1      | Depth-to-Density Conversion . . . . .                                  | 27        |
| A.2      | Evaluation Metrics: Full Definitions . . . . .                         | 27        |
| A.3      | Additional Results . . . . .   | 28        |
| A.3.1    | Baseline Modelling: Linear Regression . . . . .                        | 28        |
| A.3.2    | Model Advancements: Trend and Seasonality . . . . .                    | 28        |
| A.3.3    | Model Advancements: 1D-CNNs . . . . .                                  | 29        |
| A.3.4    | Incorporating Observational Data (RAPID) . . . . .                     | 29        |
| A.3.5    | Linear Regression Results for 30°S (ACCESS) . . . . .                  | 30        |
| A.4      | Experiment Details for Full Southern Ocean Methods . . . . .           | 30        |
| A.5      | Notes on ECCO Outputs . . . . .  | 30        |

# 1 Introduction

*Nina Baranduin*

Changes to global ocean circulations have cascading detrimental climate consequences [80]. Various locked-in debates exist over how currents are responding to climate change, how abrupt those changes could be, and what impact they could have on the climate [36, 85]. Global overturning circulations mediate the uptake, transportation and storage of atmospheric heat and carbon [22, 57]. The Meridional Overturning Current (MOC) connects the upper, mid-depth and abyssal circulation of the oceans [63]. The primary component of mid-depth circulation is the Atlantic Meridional Overturning Circulation (AMOC) [62]. As the AMOC travels north it surfaces in the North Atlantic, where salinity increases and temperature decreases, causing the dense water mass (North Atlantic Deep Water, hereafter NADW) to descend and travel southward [50]. Surfacing in the Southern Ocean, a portion mixes with the mid-depth circulation [50]. Remaining water masses form Antarctic Bottom Water (AABW), and a circumpolar deep-water current that surrounds Antarctica (known as the Antarctic Circumpolar Current, henceforth ACC) [35]. As the AABW is one of the densest water masses [2], it sinks below NADW, covering most of the ocean floor and comprising approximately one-third of the entire oceans volume [77]; therefore forming the abyssal cell of the MOC [63]. As it moves northward, AABW facilitates the transportation and deposition of oxygen and nutrients, as well as carbon uptake and storage [59].

The AMOC has been a central element of debate as it mediates the connection between hemispheres [48]; and due to the accessibility as well as climate conditions of the AMOC, sensors can be deployed for direct observations [42]. Such conditions are not present in the Southern Ocean, where it is comparatively more difficult to monitor ocean currents [43]. Conventional theory states that as the MOC changes, anti-phased temperature changes can be seen in the poles [8]. Previously, a northern freshwater influx slowed NADW and cooled the northern hemisphere, while in the southern hemisphere temperature increased. This bi-polar seesaw is thought to have occurred over decades, with a time delay in this seesaw shown in ice cores [42, 55]. However it is unknown how the bi-polar seesaw manifests in the current climate, whether the poles are in-phase or anti-phased, or how abrupt climate change may impact this polar heat transfer on known timescales [9].

As currents are driven by changes in density from differences in pressure, salinity, and temperature [32]; as well as surface winds which cause an observable change in sea surface height, it has been proposed that such satellite-observable conditions can be used to infer ocean currents [10, 11, 75]. One such paper recently claimed to have used these variables using a neural network to create a predictive model for MOC strength in the Atlantic and Southern Ocean that outperforms previous methods [71]. However these results were based using one ocean circulation model [12] and have not been replicated. It is also not known if this method outperforms more basic statistical models. Moreover, it could be possible that using direct observational MOC strength data may improve such predictive models [5, 73]. This observational data obtained from higher northern latitudes could also predict possible bi-polar seesaw response, and account for the connectedness of the polar oceans through the AMOC.

Therefore this study aims to assess how satellite-observable variables can be used to infer abyssal MOC strength; and if deep learning methods can infer latitude-invariant dependencies for the abyssal MOC strength from remote sensing variables. By replicating and improving upon previous methods [71], this is the first study to provide a baseline for predicting the abyssal MOC strength using regularised linear regression. By incorporating ACCESS [30], a high-resolution ocean circulation model, and RAPID [44], an array of sensors that directly measures AMOC strength; this is the first study to use these to improve predictive models for abyssal currents in the Southern Ocean. This study provides initial evidence that spatiotemporally aware deep learning models such as Convolutional Neural Networks (CNNs) can be reliable when used to model abyssal currents in the Southern Ocean.

## 2 Methods

*Emilio Luz-Ricca, Aline Van Driessche, Tom Cowperthwaite, & Sharan Maiya*

### 2.1 Numerical Models

*Emilio Luz-Ricca*

We use outputs from two numerical models: the “Estimating the Circulation and Climate of the Ocean” (ECCOv4r4, ECCO hereafter) ocean state estimate [14] and the “Australian Community Climate and Earth System Simulator Ocean Model, Version 2” (ACCESS-OM2-01, ACCESS hereafter) [29]. ECCO is based on the MITgcm numerical model [1] with a sea-ice module; grid cells are at a  $1^{\circ}$  resolution with 50 vertical levels. ECCO assimilates real-world oceanographic data, using MITgcm to interpolate between observations in a physically-consistent manner. Monthly ECCO outputs were retrieved from the NASA Physical Oceanography Distributed Active Archive Center (PO.DAAC) for 1992-2017. ACCESS is a global ocean-sea ice model based on MOM5 [20] with an eddy-rich  $0.1^{\circ}$  resolution and 75 vertical levels; it is able to accurately capture the physical dynamics of AABW formation and abyssal circulation [33]. Monthly ACCESS outputs were provided to us for 1989-2018.

### 2.2 Ocean Basins and Latitudes of Interest

*Emilio Luz-Ricca*

For analysing ECCO outputs, we follow [71] and split the world’s oceans into three primary basin masks: an Atlantic mask, an Indo-Pacific mask, and a Southern mask (Figure 1). The Atlantic mask includes the entire Atlantic Ocean from  $34^{\circ}\text{S}$  to  $55^{\circ}\text{N}$ , excluding the Mediterranean Sea. The Indo-Pacific mask consists of all oceans between  $34^{\circ}\text{S}$  and  $55^{\circ}\text{N}$  that are not included in the Atlantic mask. Finally, the Southern mask includes all latitudes below  $34^{\circ}\text{S}$ . In some cases, we alternatively define the ECCO Southern mask as all latitudes below  $30^{\circ}\text{S}$  to match the ACCESS outputs.

Again, we follow [71] and study ML performance at four specific latitudes in ECCO:  $26.5^{\circ}\text{N}$  in the Atlantic Ocean,  $30^{\circ}\text{S}$  in the Indo-Pacific Ocean, and  $55^{\circ}\text{S}$ ,  $60^{\circ}\text{S}$  in the Southern Ocean. For ACCESS, we examine ML model performance at  $30^{\circ}\text{S}$  and  $60^{\circ}\text{S}$  in the Southern Ocean, as these were the latitudes we were provided with.

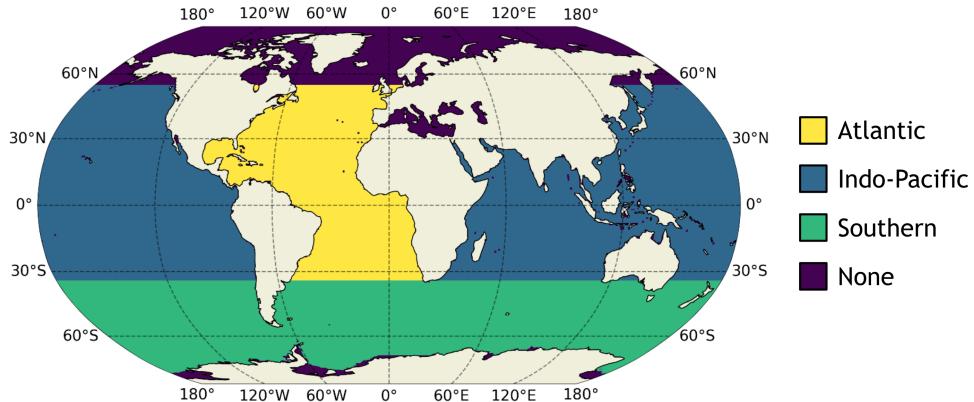


Figure 1: The three basin masks used in this study, following the mask definitions in [71].

### 2.3 Satellite-observable Variables

*Tom Cowperthwaite*

Traditional ship hydrography results in sparse measurements, and thus suffer from aliasing effects, especially temporally [15]. Advancements in satellite measurement facilitate the collection of oceanic variables on a global scale and with a high temporal resolution, therefore reducing aliasing [60]. However, satellite measurements are typically poor at resolving properties of the ocean at depth, and thus have inherent limits on usefulness when considering abyssal ocean circulations [6].

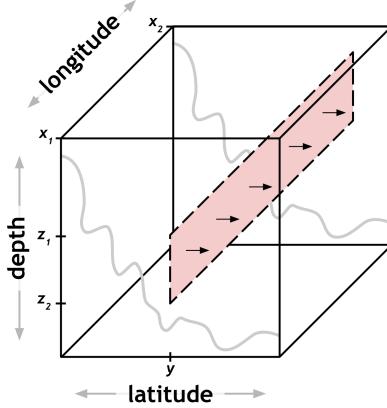


Figure 2: A sketch of MOC strength as captured by the cross-basin streamfunction  $\Psi$ . Net flow rate  $\Psi(y, t)$  is the velocity integrated over a reference surface (here as a simplified rectangular plane), defined by a latitude  $y$ , depth bounds  $(z_1, z_2)$ , and longitude bounds  $(x_1, x_2)$ , at a time  $t$ . Depth bounds are not static, but a function of latitude, longitude, and the chosen potential density  $\sigma_2$ . The seafloor is shown in gray.

It has been suggested that satellite-observable variables can provide partial information on MOC strength in deep circulation cells. Firstly, meridional Ekman transport in the upper ocean is expected to be calculable from measurements of zonal wind stress (ZWS), which may be taken via satellite [4]. Flows due to Ekman transport have been shown to be significant drivers of circulation at upper levels; they also have large impact on deep ocean currents due to the effects of Ekman suction and pumping [66]. In the Southern Ocean, the abyssal export has been shown to be highly correlated to total meridional Ekman flux, which is highly dependant on the basin-zonal-mean ZWS. This observation allows us to use basin-zonal-mean ZWS to represent the aggregate effect of wind stress on the abyssal cell transport across an entire latitude [74].

Another driver of large scale ocean currents is geostrophic flow, resulting from the approximate balance between pressure gradient and Coriolis forces [18]. This effect does not diminish with depth. Despite ocean bottom pressure (OBP) being associated with the sea floor, OBP fluctuations can cause anomalies in The Earth's gravitational field which may be detected by satellite [25, 78]. Other variables expected to play a significant role in deep water formation at high latitudes, such as sea surface temperature (SST), sea surface height (SSH), and sea surface salinity (SSS), are also available as model outputs from ECCO and ACCESS [58, 76].

In our study, we use the five variables described in this section, either zonally-resolved or zonal averages, as inputs to predictive models, as well as *in-situ* measurements of MOC strength from cross-basin arrays in Section 3.4.

## 2.4 MOC Strength Calculation

*Aline Van Driessche*

Choosing an appropriate measure of MOC strength is of utmost importance to model training and evaluation. To facilitate direct performance comparison, we follow the definition in [71] to calculate the monthly-mean MOC streamfunction  $\Psi$  in units of Sverdrups (Sv;  $1 \text{ Sv} \equiv 10^6 \text{ m}^3 \text{s}^{-1}$ ):

$$\Psi(\text{basin}, y, \sigma_2, t) = - \int_{x=x_1(\text{basin}, y)}^{x_2(\text{basin}, y)} \int_{z=\eta_b(x, y)}^{\eta_{\sigma_2}(x, y, t)} v_r(x, y, z, t) dz dx \quad (1)$$

Here,  $x$  denotes longitude,  $y$  is the latitude of interest, and “basin” refers to a basin mask (Section 2.2) with  $x_1$  and  $x_2$  as the zonal basin boundaries. A sketch of the physical property that the streamfunction captures is shown in Figure 2. Potential density ( $\sigma_2$ ), bottom depth ( $\eta_b$ ), isopycnal depth ( $\eta_{\sigma_2}$ ), and residual velocity ( $v_r$ ) will be detailed subsequently.

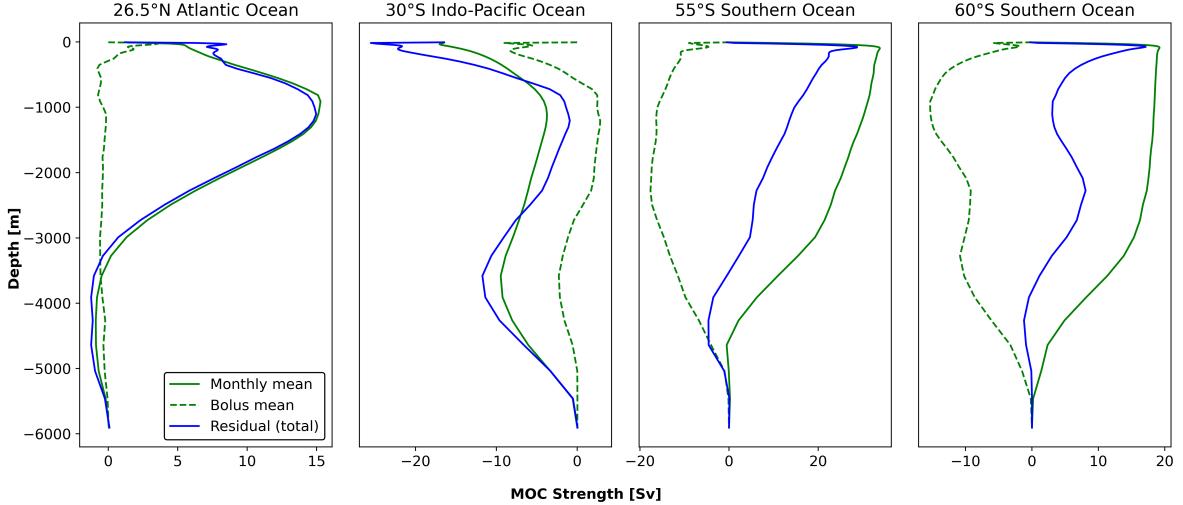


Figure 3: Time averaged MOC strength in depth space calculated with the monthly mean, bolus and residual (sum) velocities for latitudes  $26.5^{\circ}\text{N}$ ,  $30^{\circ}\text{S}$ ,  $55^{\circ}\text{S}$  and  $60^{\circ}\text{S}$ .

The residual velocity  $v_r = v_m + v_e$  contains two components: the monthly mean (Eulerian) velocity  $v_m$  and the “bolus” velocity  $v_e$ , a parameterized eddy-induced monthly mean velocity that uses an artificial velocity field to mimic the advective transport by eddies (local circular currents) [17]. Interaction between strong westerly winds and the ACC makes the bolus component significantly stronger in the Southern Ocean [3] ( $55^{\circ}\text{S}$  and  $60^{\circ}\text{S}$  in Figure 3).

We evaluate the MOC strength  $\Psi$  at each timestep at the density level  $\sigma_2$  at which the time-mean streamfunction is maximal [15], defining  $\sigma_2^*$  as:

$$\sigma_2^*(\text{basin}, y) = \operatorname{argmax}_{\sigma_2} \left\{ \left| \overline{\Psi}(\text{basin}, y, \sigma_2, t) \right| \right\}, \quad (2)$$

where  $\overline{\Psi}(\cdot)$  represents the time-mean, density-space streamfunction (details in Appendix A.1). We maximise over a subset of potential densities based on the cell of interest; for the abyssal cell, this corresponds to maximal northwards flow.

In oceanography, an isopycnal is a level at which density remains constant. Figure 4 demonstrates that isopycnals vary across depths and temporally. This underscores the complexity of ocean stratification: a singular density value can correspond to multiple depth levels over time.

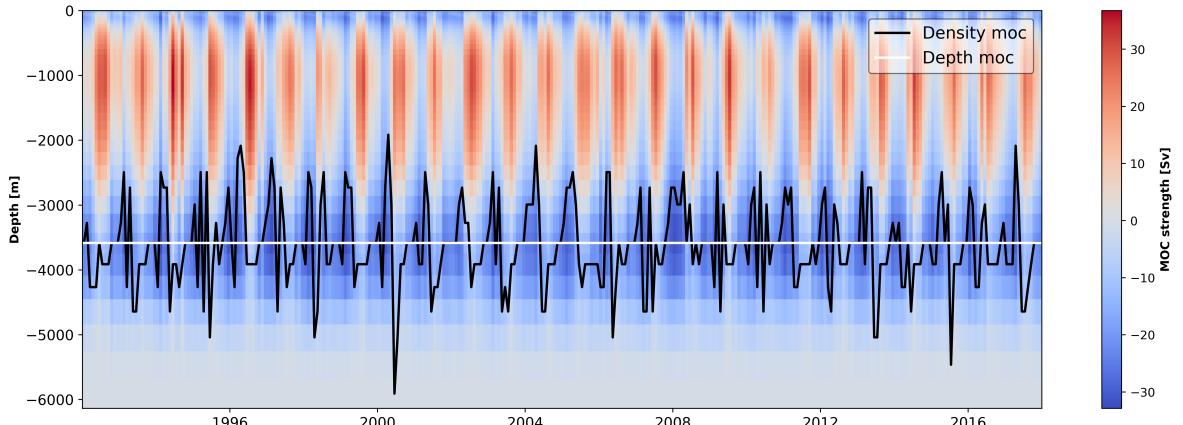


Figure 4: Streamfunction at  $30^{\circ}\text{S}$  in time- and depth-space. The white line indicates the depth where the time-mean MOC is maximal and the black line indicates the depths corresponding to density level  $\eta_{\sigma_2}$  at which the time-mean MOC is maximal, averaged zonally.

[71] calculates the MOC strength in the native lat-lon-cap 90 grid (LLC90). This grid decomposes the Earth into 13 tiles, each  $90 \times 90$  with 50 depth layers [14, 84] (Figure 5). However practical for representation, the different tile orientations complicate tasks such as basin masking. Therefore, we compute the streamfunction in an interpolated latitude-longitude grid which generates errors, but generally they are minor (Figure 6).

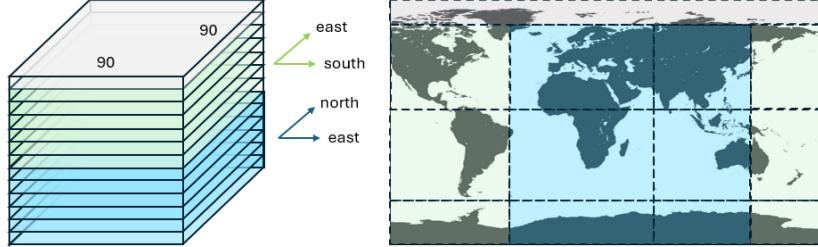


Figure 5: The LLC90 model grid representation. The Earth splits into 13 stacked tiles; 6 designate the  $+x$  and  $+y$  directions as North-East, 6 as East-South, and 1 tile represents the North Pole.

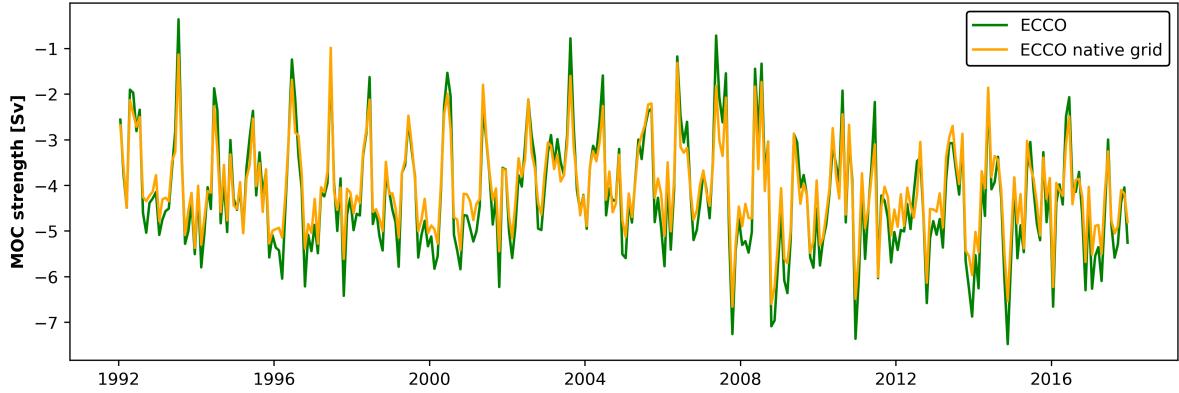


Figure 6: MOC strength in the (Atlantic) abyssal cell at  $30^\circ\text{S}$  calculated in the native model grid vs. an interpolated latitude-longitude grid.

We apply the same streamfunction definition to ACCESS. As we were not provided the potential temperature information needed to perform calculations in density-space, we use depth-space instead. While the ACCESS and ECCO MOC strength timeseries do not align perfectly, they have the same order of magnitude and show similar trends (Figure 7).

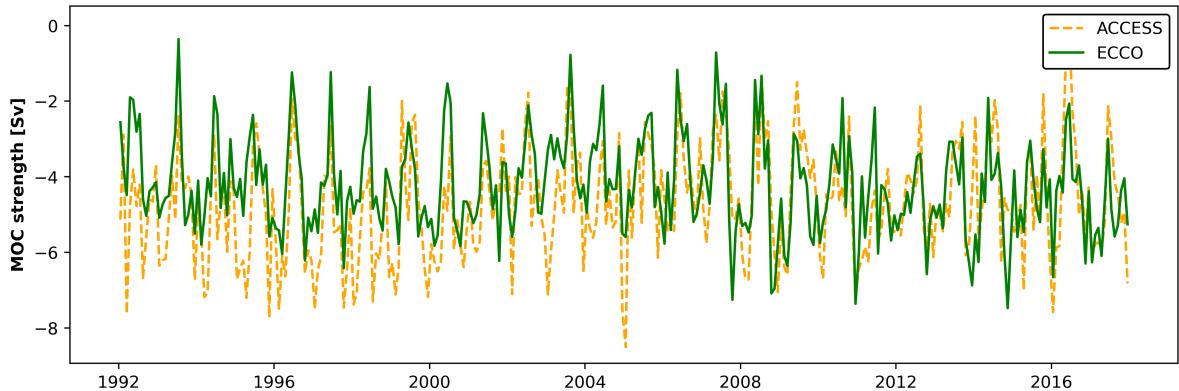


Figure 7: MOC strength for the abyssal cell at  $30^\circ\text{S}$  in the Atlantic (depth space) for ECCO and ACCESS.

## 2.5 Variable Preprocessing

*Emilio Luz-Ricca*

We standardize inputs by removing the series time-mean and dividing by the time-standard deviation independently for each longitude and each satellite-observable variable. In some cases, we also remove the time trend and seasonality from the MOC time series using an additive decomposition:

$$\Psi_t = \text{trend}(\Psi_t) + \text{seasonal}(\Psi_t) + \text{residual}(\Psi_t),$$

where  $\Psi_t$  represents the ECCO MOC strength at time  $t$ . Whenever we detrend and deseasonalize the MOC time series, we also detrend and deseasonalize the series for satellite-observable input variables. For some ML models, we also take zonal averages of inputs after preprocessing.

## 2.6 Machine Learning Training Details

*Sharan Maiya*

Existing applications of ML to similar problems include [71], which to our knowledge constitutes the current state-of-the-art. The authors train linear and nonlinear neural networks to infer MOC strength at different latitudes, focusing particularly on 26.5°N, 30°S, 55°S, and 60°S. In Section 3 we establish baseline results with linear regression before replicating the methodology in [71].

From here, we explore more complex architectures to leverage spatial and temporal information; these include convolutional neural networks (CNNs) [31] and sequence-to-sequence architectures such as the recurrent neural network (RNN), gated recurrent unit (GRU) [7], and long short-term memory (LSTM) [23]. For an overview of these architectures and deep learning practices see [19]. All our model implementations can be found on our GitHub repository (see Section 5).

We follow standard ML practices for feature selection and hyperparameter tuning, using either a train/test split with  $k$ -fold cross-validation or a train/validation/test split in the case of computationally intensive training runs. All models are trained on consumer hardware utilising GPU acceleration where possible. This is carried out through the PyTorch [53] deep learning framework.

## 2.7 Evaluation Metrics

*Emilio Luz-Ricca*

We use four standard regression metrics to evaluate ML model performance: mean absolute error (MAE), root mean squared error (RMSE), (thresholded) mean absolute percentage error ( $\text{MAPE}_\tau$ ), and (linear) correlation ( $r$ ). See Appendix A.2 for the full definitions and interpretation of each metric. In tables, we use  $\uparrow$  ( $\downarrow$ ) to remind readers that higher (lower) values indicate better performance for a particular metric.

# 3 Results & Discussion

*All Authors*

Our analysis begins with a replication of the approach taken in [71] and a comparison to a standard baseline method (linear regression). We then extend this work by training a suite of new models which are better able to predict residual MOC strength, including its trend and monthly seasonal components, and discover an apparent approximately linear relationship.

Finally, we present three novel experiments:

1. Use of observational data at northern latitudes (RAPID) to improve model performance at southern latitudes,
2. Generalisability to all latitudes in the Southern Ocean by learning latitude-invariant features,
3. Use of high-quality simulation outputs (ACCESS) as ground truth to validate model reliability.

In contrast to [71], we focus on the abyssal MOC cell in all settings.

### 3.1 Baseline Modelling: Linear Regression

*Emilio Luz-Ricca*

To establish a predictive baseline, we thoroughly test regularised linear regression using ElasticNet [89] for predicting ECCO density-space MOC strength; we use implementations provided by **Scikit-Learn** [54] and **statsmodels** [67]. We inspect model performance with different combinations of inputs along three axes:

- Longitudes: zonal averages of satellite-observable variables vs. full zonal information for each variable,
- Time: static in time (i.e., using only variables for the current time step) vs. variable history (i.e., providing data for past months for each variable),
- Satellite-observable variables: all variables (i.e., SSH, SST, SSS, OBP, ZWS) vs. some subset of variables.

In all linear regression experiments, we train on the first 70% and test on the last 30% of the time series. We evaluate model performance for detrended/deseasonalized inputs and outputs to match [71], as well as the raw MOC time series (trend/seasonality retained). We focus on 60°S here (Tables 1 and 2), but full results for other latitudes can be found in the Appendix (Tables 12 and 13).

In general, we find that linear regression performs very well across the latitudes of interest (Figure 8). At 60°S, the best performing models explain more than 95% of the time-variance in MOC strength (Tables 1 and 2). While [71] do not present full results for linear regression, we find that it provides a much more competitive baseline than they report. This is evidence that the relationship between satellite-observable variables and abyssal circulation, while undoubtedly complex, is approximately linear.

| Inputs                            | MAE $\downarrow$ | RMSE $\downarrow$ | MAPE $_{0\downarrow}$ | MAPE $_{0.5\downarrow}$ | Correlation $\uparrow$ | $R^2\uparrow$ |
|-----------------------------------|------------------|-------------------|-----------------------|-------------------------|------------------------|---------------|
| all inputs, zonal avg.            | 2.70             | 3.58              | 108.93%               | 88.47%                  | 0.87                   | 0.74          |
| OBP + SSH + ZWS, full zonal       | 2.04             | 2.45              | 103.68%               | 63.61%                  | 0.96                   | 0.88          |
| OBP, full zonal                   | <b>1.05</b>      | <b>1.39</b>       | <u>57.58%</u>         | <b>35.34%</b>           | <b>0.98</b>            | <b>0.96</b>   |
| OBP, full zonal + ZWS, zonal avg. | <u>1.10</u>      | <u>1.45</u>       | <b>53.34%</b>         | 37.92%                  | <b>0.98</b>            | <b>0.96</b>   |

Table 1: Performance for ElasticNet on 60°S. Model inputs and outputs *are* detrended and deseasonalized. Best results are in bold and second best are underlined.

| Inputs                                   | MAE $\downarrow$ | RMSE $\downarrow$ | MAPE $_{0\downarrow}$ | MAPE $_{0.5\downarrow}$ | Correlation $\uparrow$ | $R^2\uparrow$ |
|--|------------------|-------------------|-----------------------|-------------------------|------------------------|---------------|
| all inputs, zonal avg.                   | 4.59             | 5.37              | 101.31%               | 68.56%                  | 0.88                   | 0.46          |
| all inputs, 6-month history, zonal avg.  | 2.55             | 3.26              | 69.73%                | 49.76%                  | 0.90                   | 0.80          |
| all inputs, 12-month history, zonal avg. | 2.64             | 3.43              | 74.96%                | 56.28%                  | 0.89                   | 0.78          |
| OBP + SSH + ZWS, full zonal              | 1.42             | 1.81              | 39.57%                | 30.44%                  | <u>0.98</u>            | <u>0.94</u>   |
| OBP, full zonal                          | <u>1.32</u>      | <u>1.70</u>       | <u>35.87%</u>         | <u>25.73%</u>           | <u>0.98</u>            | <b>0.95</b>   |
| OBP, full zonal + ZWS, zonal avg.        | <b>1.25</b>      | <b>1.60</b>       | <b>34.68%</b>         | <b>24.99%</b>           | <b>0.99</b>            | <b>0.95</b>   |

Table 2: Performance for ElasticNet on 60°S. Model inputs and outputs *are not* detrended and deseasonalized. Best results are in bold and second best are underlined.

### 3.2 Replication of Solodoch *et al.*

*Sharan Maiya*

Having established baseline results with linear regression, we now explore deep learning, beginning with a replication of the methodology in [71].

We model MOC strength using only full zonal information of OBP. This choice, and all other implementation details, are kept consistent with the top performing models in [71] except where explicitly specified.

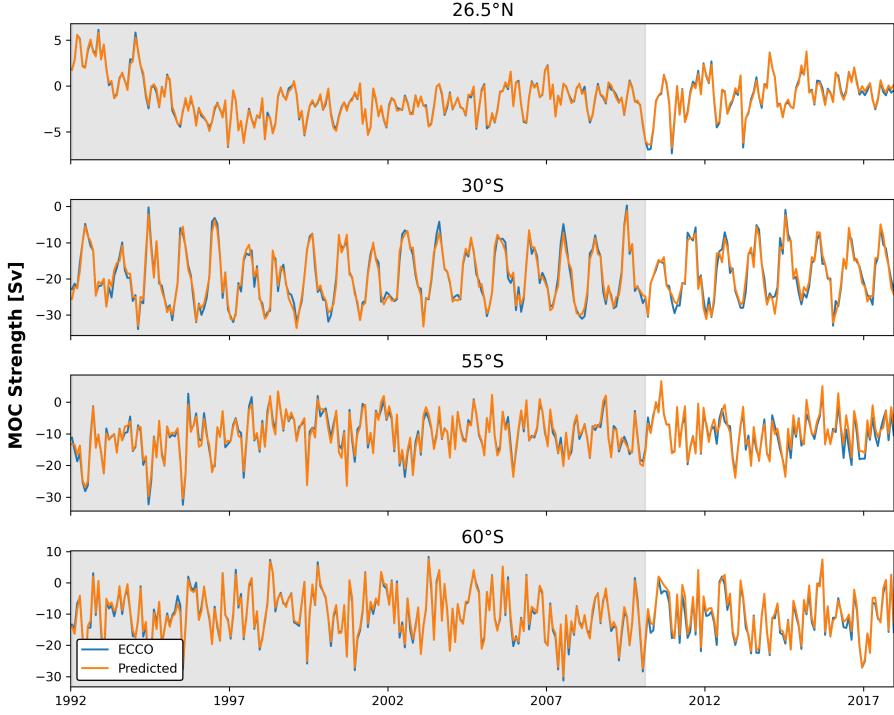


Figure 8: ElasticNet model predictions (inputs: OBP, full zonal + ZWS, zonal avg.) vs. ECCO MOC strength (trend/seasonality retained). Training data are plotted against a grey background.

### 3.2.1 Training Details

*Sharan Maiya*

We found two significant issues with the approach in [71] after inspecting the code archive [72]:

1. Data leakage occurs due to the calculation of trend, seasonality, and feature scaling parameters over the entire dataset. We address these issues in our preprocessing.
2. The implemented method of feature scaling (MinMax) differs from the stated method in the paper (Z-Score). We follow the method described in the paper.

Our model is an MLP [65] with a single hidden layer of one neuron, using a ReLU [51] activation. While the model in [71] was trained via Bayesian Interpolation [13, 40], we use batched gradient descent with AdamW [37], minimising MSE, with a learning rate of  $10^{-3}$  and  $L_2$  regularisation ( $10^{-5}$ ). We use a 60%/20%/20% training/validation/testing split with early stopping.

### 3.2.2 Model Performance

*Sharan Maiya*

| Latitude | RMSE $\downarrow$ | MAPE $_{0.1}\downarrow$ | MAPE $_{0.5}\downarrow$ |
|----------|-------------------|-------------------------|-------------------------|
| 26.5°N   | 0.29              | 7.91%                   | 7.91%                   |
| 30°S     | 2.23              | 549.90%                 | 115.53%                 |
| 55°S     | 3.28              | 148.66%                 | 80.72%                  |
| 60°S     | 2.84              | 1125.61%                | 78.02%                  |

Table 3: Test set performance for our models.

In Section 4.1 of [71] the authors report RMSE values ranging between 0.55 and 1.4. We do not recover comparable results, as shown in Table 3. While we feel the results in [71] are unreliable due to data leakage concerns discussed earlier, our data are sourced from ECCOv4r4 which includes two extra years

from ECCOv4r3 used in [71]. Notably, our test set covers these two years, which may contribute to our differing results.

Model performance is different across the four latitudes, suggesting either that the dependencies between OBP and MOC strength are spatially independent, or that any latitude-invariant features have not been learned. In Section 3.3 we explore more flexible architectures and in Section 3.5 we directly learn a model which generalises to different latitudes.

Figure 9 plots predictions for each of the models in Table 3. The best performing model (at  $26.5^{\circ}\text{N}$ ) mostly struggles to infer peak values. For the model at  $30^{\circ}\text{S}$ , we believe poor test performance is largely due to the removal of a strong seasonal signal, which can be seen in Figure 10. We discuss the difficulties of predicting at  $55^{\circ}\text{S}$  and  $60^{\circ}\text{S}$  in Section 3.5.

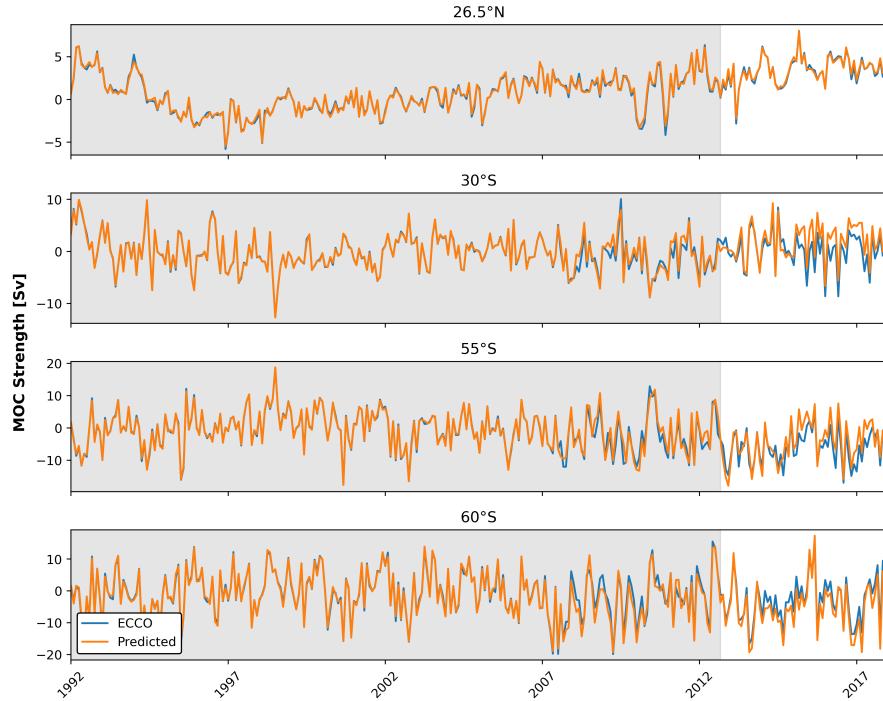


Figure 9: Predictions vs MOC strength for all models. Training and validation data are plotted against a grey background.

### 3.2.3 Limitation: Loss of Important Signal

*Sharan Maiya*

We feel that important climate risk scenarios with respect to ocean circulation include significant changes in trend or seasonality [34], both of which are removed for the above models. When included, we find they underfit: all four undergo mode collapse, as shown in Figure 10. In the next section we explore more flexible architectures and demonstrate their ability to better capture trend and seasonality instead of residual MOC strength alone.

## 3.3 Model Advancements

*Aline Van Driessche & Nina Baranduin*

Building further on the architecture of [71], we aim to improve the prediction performance through neural network (NN) optimization. We do so, as we try to find out if deep learning approaches perform above and beyond a baseline statistical model for satellite-observable based MOC predictions. First we evaluate how trend and seasonality influence CNN performance variability. For further model training, unlike earlier approaches, we do not remove the trend nor seasonal component from the MOC strength time series to retain the series’ complex characteristics. We assess NN effectiveness using both 1D- and 2D-CNNs.

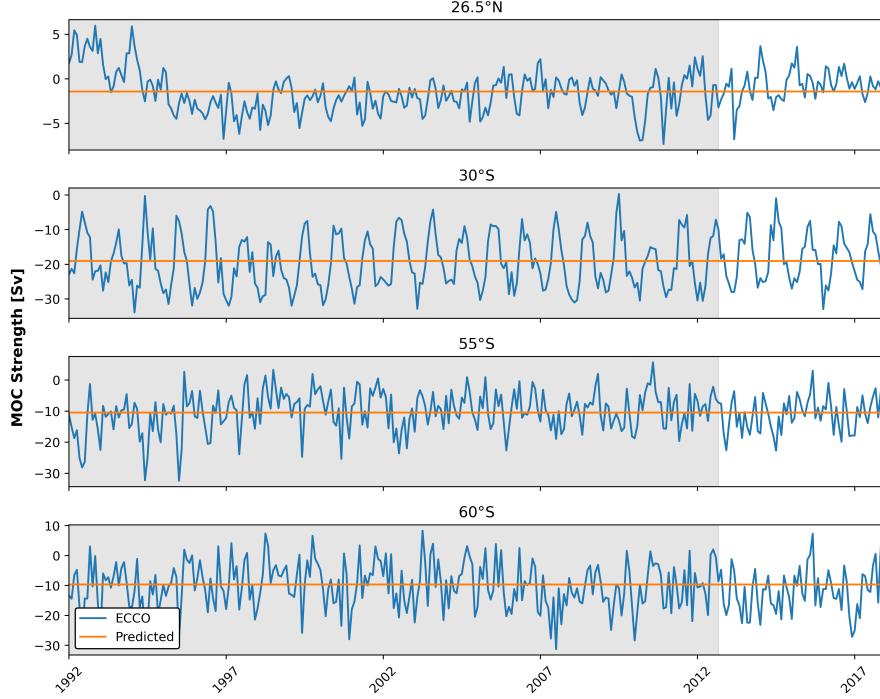


Figure 10: As in Figure 9 - models are trained **without** the removal of trend and seasonal components.

### 3.3.1 Trend and seasonality

*Nina Baranduin*

Previous models [71] using ECCO [12] removed seasonal and long-term trends in the data before training and testing models; as this can remove potentially irrelevant data therefore improve model performance [83, 87]. To properly assess how trends influence model performance, each latitude was compared separately. The results (Table 4) show that for three latitudes completely removing trends produced lower MSE, RMSE and MAE compared to when trends were included; indicating improved model performance with no trends. At latitude 26.5°N there was a 51.5%, 30.7% and 32.1% decrease in MSE, RMSE and MAE (respectively) when all trends were removed, compared to all trends included. This decrease was also shown at Latitude 30°S (260.1% MSE, 89.8% RMSE and 113.4% MAE) and latitude 55°S (276.7% MSE, 4.1% RMSE and 111.4% MAE). However including trends reduced model performance at latitude 60°S, with an increase of 12.7% (MSE), 6.6% (RMSE) and 4.26% (MAE) compared to when all trends were removed. Although this percentage increase is smaller than the percentage decrease seen at other latitudes, suggesting the inclusion of trends only marginally decreased model performance. Evaluating model performance using  $\text{MAPE}_0$  showed the opposite; all models with season and trends performed better when compared to models with trends removed. All  $\text{MAPE}_0$  scores were >50% when all trends were removed, indicating inaccurate forecasting [46, 52]. For all latitudes apart from 60°S, including trends produced the lowest  $\text{MAPE}_0$  from every comparison configuration.

As MSE, RMSE and MAE reflect absolute errors, [24, 68] if a dataset contains a small range of values these errors will reflect that scale [27]. However this would not impact relative errors, which could account for the higher  $\text{MAPE}_0$  and lower absolute error found [24]. Inconsistent errors in variance (heteroscedasticity) could also result in larger relative errors but not inflate absolute errors, increasing  $\text{MAPE}_0$  but not MSE, RMSE or MAE [81]. When season and trend are included, extreme outliers can disproportionately inflate the  $\text{MAPE}_0$  [28]. This is one of the main justifications for removing trends and also a limitation of using the  $\text{MAPE}_0$  metric [45]. Yet the low  $\text{MAPE}_0$  scores shown with trends included, suggest outliers in this dataset had little impact. Although seasonality and trends could introduce potential noise; including trends can improve the ability to capture underlying structure in the data [49, 79]. As it is unlikely outliers negatively impacted model performance, and understanding seasonal and long-term trend in the MOC strength is important for predicting abrupt changes [34, 41, 88], data should therefore include seasonality and trend henceforth.

| Latitude 26.5°N |       |             |             |             |                     |
|-----------------|-------|-------------|-------------|-------------|---------------------|
| Season          | Trend | MSE↓        | RMSE↓       | MAE↓        | MAPE <sub>0</sub> ↓ |
| NO              | YES   | 15.49       | 3.94        | 3.39        | 31.24%              |
| NO              | NO    | 6.43        | 2.54        | 1.99        | 589.35%             |
| YES             | YES   | 10.89       | 3.31        | 2.63        | <b>27.44%</b>       |
| YES             | NO    | <b>6.38</b> | <b>2.53</b> | <b>1.95</b> | 324.60%             |
| Latitude 30°S   |       |             |             |             |                     |
| Season          | Trend | MSE↓        | RMSE↓       | MAE↓        | MAPE <sub>0</sub> ↓ |
| NO              | YES   | 4.13        | 2.03        | 1.58        | 50.9%               |
| NO              | NO    | 4.18        | 2.04        | 1.58        | 54.16%              |
| YES             | YES   | 15.05       | 3.88        | 3.36        | <b>18.92%</b>       |
| YES             | NO    | <b>3.28</b> | <b>1.81</b> | <b>1.40</b> | 34.88%              |
| Latitude 55°S   |       |             |             |             |                     |
| Season          | Trend | MSE↓        | RMSE↓       | MAE↓        | MAPE <sub>0</sub> ↓ |
| NO              | YES   | <b>7.26</b> | <b>2.69</b> | 2.14        | 149.66%             |
| NO              | NO    | 7.40        | 2.72        | <b>2.12</b> | 135.45%             |
| YES             | YES   | 27.86       | 5.28        | 4.47        | <b>83.11%</b>       |
| YES             | NO    | 7.60        | 2.76        | <b>2.12</b> | 152.70%             |
| Latitude 60°S   |       |             |             |             |                     |
| Season          | Trend | MSE↓        | RMSE↓       | MAE↓        | MAPE <sub>0</sub> ↓ |
| NO              | YES   | 24.61       | 4.96        | 4.27        | <b>63.60%</b>       |
| NO              | NO    | 7.40        | 2.72        | 2.04        | 102.84%             |
| YES             | YES   | <b>6.45</b> | <b>2.54</b> | <b>1.99</b> | 92.93%              |
| YES             | NO    | 6.64        | 2.58        | 2.01        | 88.15%              |

Table 4: Performance of neural networks when removing (NO) or including (YES) seasonality and trends. Model performance is assessed by Mean Square Error (MSE), Root Mean Square Error (RMSE), Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE<sub>0</sub>). The hyper-parameters used are included in Appendix A.3.2.

### 3.3.2 1D-CNNs

*Aline Van Driessche*

Testing various satellite observable inputs and hyperparameter tuning is initially performed on latitude 55°S, as summarised in Table 5. A CNN (details in Appendix A.3) using only OBP and ZWS performs clearly best, outperforming models using the full range of inputs.

| Inputs                 | history | MAE↓        | RMSE↓       | Correlation↑ | MAPE <sub>0</sub> ↓ | MAPE <sub>0.5</sub> ↓ |
|------------------------|---------|-------------|-------------|--------------|---------------------|-----------------------|
| all inputs, zonal avg. | 1 month | 3.45        | 4.23        | 0.69         | 69.34%              | 51.56%                |
| all inputs, zonal avg. | 6 month | 4.82        | 6.04        | 0.45         | 65.20%              | 65.20%                |
| OBP + ZWS, full zonal  | -       | 3.44        | 4.18        | 0.89         | 89.00%              | 66.30%                |
| OBP + ZWS, zonal avg.  | 1 month | <b>3.36</b> | <b>4.06</b> | <b>0.71</b>  | <b>42.93%</b>       | <b>42.93%</b>         |
| OBP + ZWS, zonal avg.  | 6 month | 4.06        | 4.89        | 0.65         | 96.32%              | 82.94%                |

Table 5: Performance of 1D-CNNs on 55°S in density space using different input variables, with and without historic data (in months).

We continue leveraging OBP and ZWS as input variables on other latitudes. Adjusting network depths and kernel sizes for individual latitudes yields slightly better results. The results in Table 6 indicate that, for latitudes further in the Southern Ocean (55°S, 60°S), model performance improves with more months of data included. Meanwhile, for 30°S this leads to accelerated over-fitting, reducing generalisability on the test data (Figure 11).

| Latitude | history  | d | kernel | RMSE $\downarrow$ | MAPE $_{0.5}\downarrow$ | MAPE $_{0.5}\downarrow$ |
|----------|----------|---|--------|-------------------|-------------------------|-------------------------|
| 26.5°N   | 1 month  | 2 | 5      | <b>2.21</b>       | 97.50%                  | 97.50%                  |
| 30°S     | 1 month  | 3 | 5      | 2.90              | <b>16.50%</b>           | <b>16.50%</b>           |
| 55°S     | 6 months | 3 | 3      | 4.07              | 42.93%                  | 42.93%                  |
| 60°S     | 6 months | 3 | 3      | 5.26              | 43.40%                  | 43.40%                  |

Table 6: Performance of 1D-CNNs across latitudes 26.5°N, 30°S, 55°S and 60°S with optimised amount of historical data (months) included.

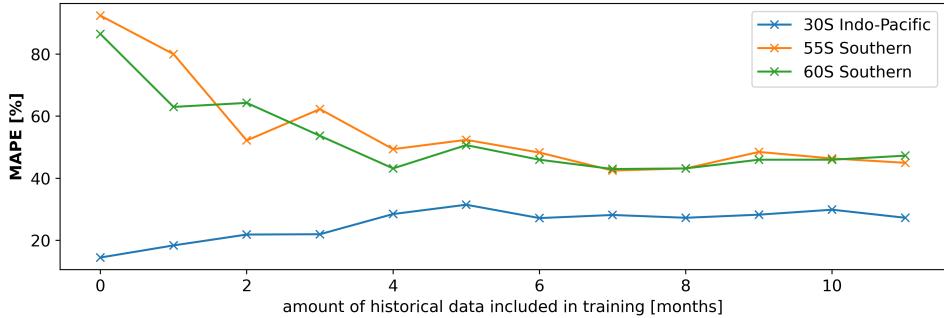


Figure 11: MAPE performance for 1D-CNNs with optimised hyperparameters per latitude, including historical data up to 12 months per timestep.

As indicated earlier in Section 3.2, certain latitudes are easier to fit (visualisation in Appendix A.3). 30°S consistently performs quite well, presumably due to its strong seasonal component (i.e. from seasonal winds) [56], generating a stable pattern. Conversely, at latitudes such as 26.5°N, considerable MOC strength fluctuations over time complicate accurate prediction. Potentially because its small abyssal component makes it prone to upper-layer highly variable processes [26]. At latitudes in the Southern Ocean (55°S and 60°S), complex ocean interactions such as the ACC introduce irregularity [64]: the models approximate the overarching trend relatively well, yet struggle to accurately capture MOC strength peak values.

### 3.3.3 2D-CNNs

#### Aline Van Driessche

To include both time and longitude variability we use 2D-CNNs. For each latitude, we extend the best performing 1D-CNN to a 2D model, using the most informative inputs (OBP and ZWS). The performances in Table 7 reveal a similar pattern on the use of historical data: more insightful for the southern latitudes (55°S and 60°S) than the others (30°S and 26.5°N).

| Latitude | history  | d | kernel | RMSE $\downarrow$ | MAPE $_{0.5}\downarrow$ | MAPE $_{0.5}\downarrow$ |
|----------|----------|---|--------|-------------------|-------------------------|-------------------------|
| 26.5°N   | 1 month  | 2 | 5      | <b>1.58</b>       | 154.80%                 | 94.70%                  |
| 30°S     | 1 month  | 2 | 3      | 2.47              | <b>12.90%</b>           | <b>12.90%</b>           |
| 55°S     | 6 months | 2 | 3      | 4.94              | 42.90%                  | 42.90%                  |
| 60°S     | 6 months | 3 | 5      | 5.26              | 43.40%                  | 43.40%                  |

Table 7: Performance of 2D-CNNs across latitudes 26.5°N, 30°S, 55°S and 60°S with optimised amount of historical data (months) and optimised network depths (d) and kernel sizes.

At 55°S, the MAPE $_{0.5}$  indicates improved model performance with historical data, yet visual inspection of the time series contrasts that (Figure 12). This suggests we should be cautious with relying solely on performance metrics as they might aggregate important information. The time-static model presents a time series that resembles more realistic fluctuations but it encounters area-specific difficulties,

as evidenced by the higher  $MAPE_{0.5}$ . Conversely, the model with history only generalizes the trend, overlooking nuances and more detailed variations in MOC strength.

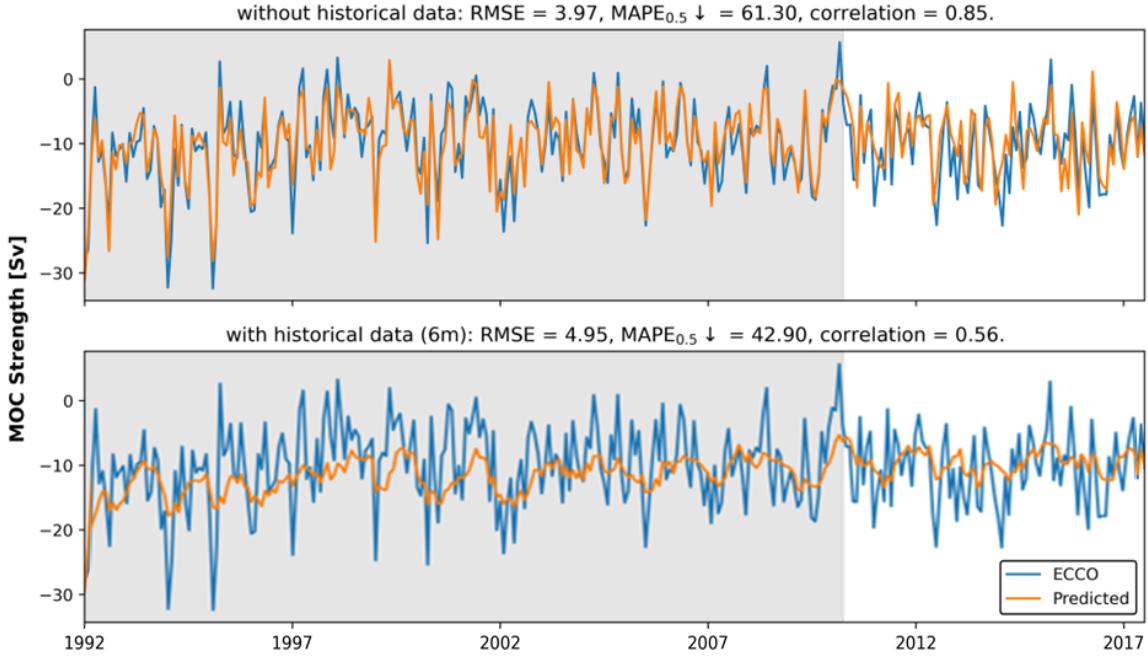


Figure 12: Predictions vs true MOC strength value for 55°S. Comparison between a 2D-CNN model (over time and longitude) with and without historical data (6 months) input.

Finally, we compare the best performing network across all latitudes, highlighting that, unlike in Section 3.2, these networks are trained on MOC strength with linear trend and seasonal components. Table 8 demonstrates that our models surpass the performance of the initial [71] architecture, yet more advanced architectures, such as 1D- and 2D-CNNs, do not yield substantial improvements.

| Latitude | Linear Regression | MLP ([71]) | MLP (ours) | 1D-CNN | 2D-CNN  |
|----------|-------------------|------------|------------|--------|---------|
| 26.5°N   | <b>43.68%</b>     | 229.82%    | 81.25%     | 97.50% | 154.80% |
| 30°S     | <b>9.96%</b>      | 75.84%     | 22.40%     | 12.90% | 12.90%  |
| 55°S     | <b>26.81%</b>     | 73.17%     | 37.98%     | 35.00% | 42.90%  |
| 60°S     | <b>34.68%</b>     | 187.24%    | 45.79%     | 43.10% | 43.40%  |

Table 8: Summary of top model performance ( $MAPE_0$ , lower is better) for abyssal MOC strength prediction at all four latitudes studied. Model performance is reported on MOC strength *including* linear trend and monthly seasonal components, to provide comparable results.

Subsequently we aim to improve on the reported performances through three novel experiments: incorporating observational data, generalisation to all latitudes in the Southern Ocean and using high-quality simulation outputs (ACCESS).

### 3.4 Experiment 1: Incorporating Observational Data (RAPID)

*Aline Van Driessche*

Observational MOC strength measurements can be available from *in-situ* arrays like the RAPID-MOCHA array, continuously recording the MOC strength at 26.5° (across the Atlantic) since 2004 [47]. We use this to assess if adding observational data from northern latitudes improves the predictability of abyssal MOC strength at southern latitudes. RAPID information is only available in depth-space, thus we train

on MOC strength in depth-space (as opposed to earlier sections). Our analysis focuses on the Atlantic mask at  $30^{\circ}\text{S}$ , leveraging regional mass conservation, bounded by South America and Africa [70].

Using the linear regression setup from Section 3.1 with ZWS and 1D-CNN architecture from Section A.3.3 with OBP and ZWS yields the best reference performances without RAPID. We compare this with models including RAPID data from several months earlier than the  $30^{\circ}\text{S}$  MOC strength at each timestep, to account for the transport delay of water mass across latitudes [61]. Figure 13 (representing average results over 100 random train-test splits) indicates that the satellite-observable variables have a stronger predictive power than RAPID—the models using RAPID data rarely outperform the baseline model. The 1D-CNN results are less conclusive, likely due to insufficient data for substantial learning, as the topmost bar plot in Figure 13 shows. However, experiments on using RAPID alone suggest that there might be a transfer of information across both depth and distance from  $26^{\circ}\text{N}$  to  $30^{\circ}\text{S}$ , especially when combining multiple time lags (see Appendix A.3). However, these findings, which hint at a bipolar seesaw effect, are not robust and need further verification to draw conclusions on the relationship between observational data and abyssal MOC strength.

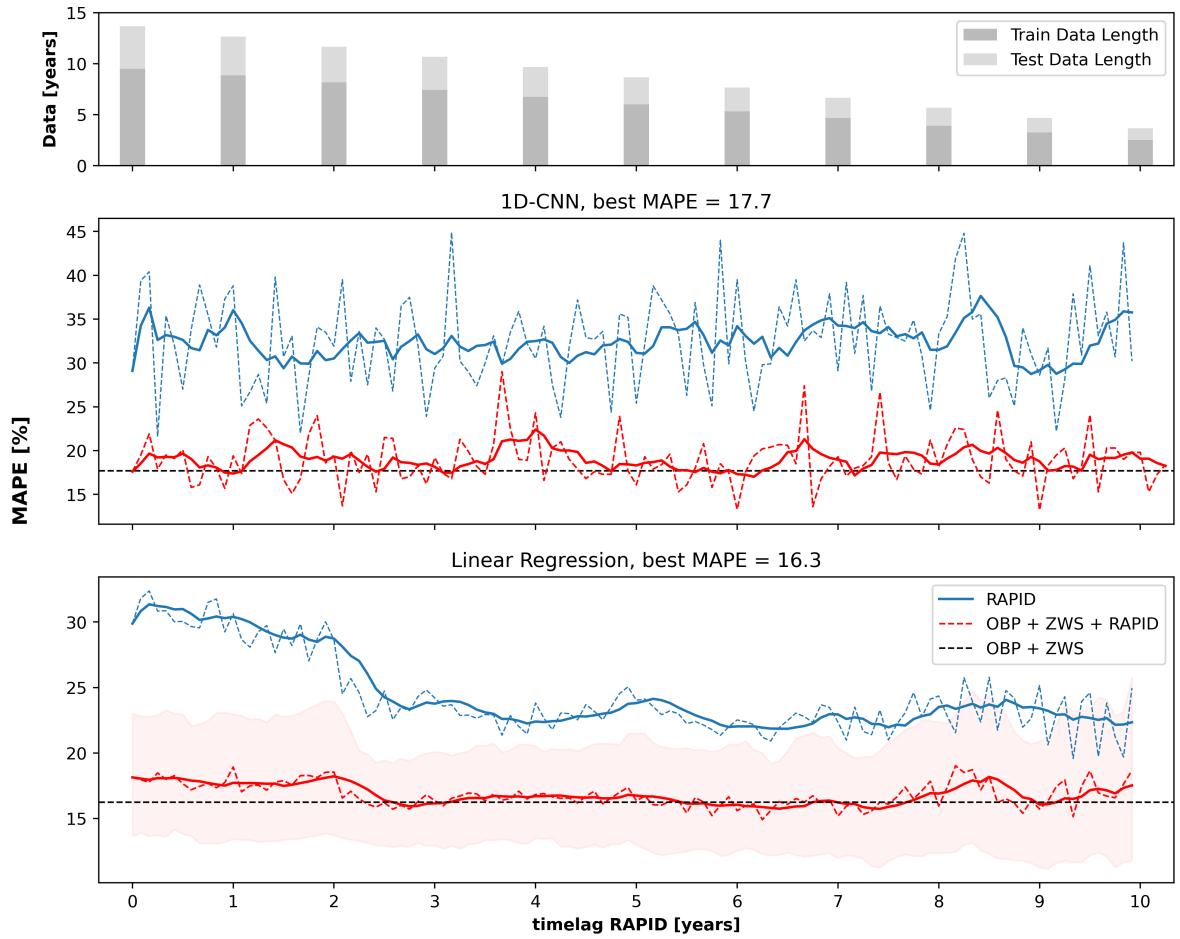


Figure 13: Timelags of RAPID used for training versus MAPE error when predicting at  $30^{\circ}\text{S}$ .

### 3.5 Experiment 2: Modelling All Latitudes in the Southern Ocean

*Sharan Maiya & Emilio Luz-Ricca*

Thus far we have studied four specific latitudes, but we now address the question: are the patterns learned by our models latitude-dependent or can they learn latitude-invariant relationships? We investigate this question by studying MOC strength across the entire Southern Ocean.

### 3.5.1 Linear Regression

*Emilio Luz-Ricca*

Following the impressive results in Section 3.1, we first experiment with ElasticNet model transfer. To do this, we train on only 60°S using zonal averages of all input variables and then use the trained model to predict the entire MOC time series at all Southern Ocean latitudes. Only 80.5°S to 30°S are considered, because latitudes below 80.5°S intersect Antarctica. Surprisingly,  $\text{MAPE}_0$  is consistently lower at most other latitudes, with an increase in  $\text{MAPE}_0$  around 60°S (Figure 14), which indicates the MOC strength is less predictable in this section.

We then train and test an independent ElasticNet model for each latitude in the Southern Ocean, with full zonal OBP as the only model input. Again, we find relatively high  $\text{MAPE}_0$  around 60°S (Figure 15), providing further evidence for the apparent difficulty of this section and motivating the use of more advanced predictive methods.

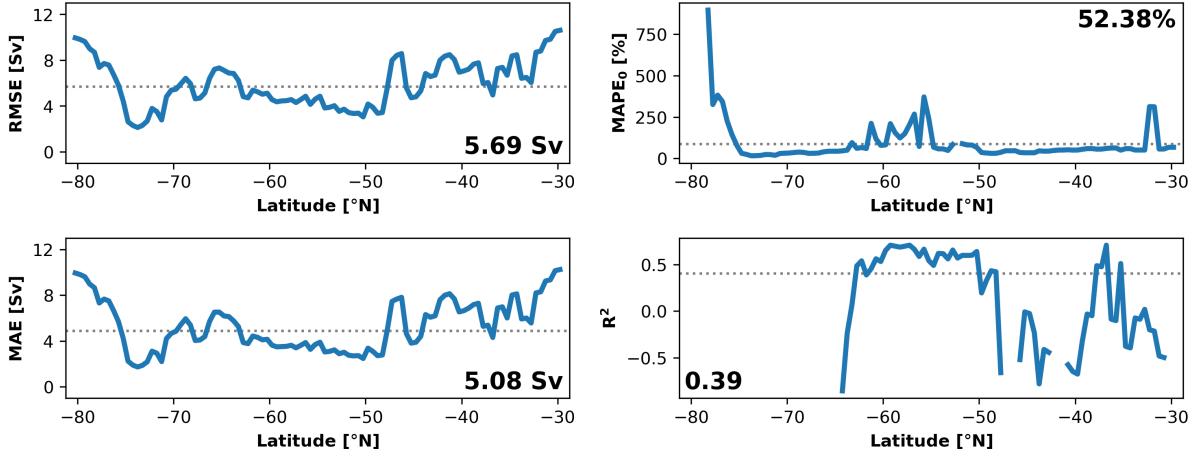


Figure 14: Generalisability of ElasticNet trained only on 60°S to all Southern Ocean latitudes (80.5°S to 30°S). The dotted line indicates the test performance of the model on the test set for 60°S and the bold inset is the median performance across all latitudes.  $\text{MAPE}_0$  values greater than 1000% and  $R^2$  values less than -1 are removed from the plot (and median calculations) to facilitate interpretation.

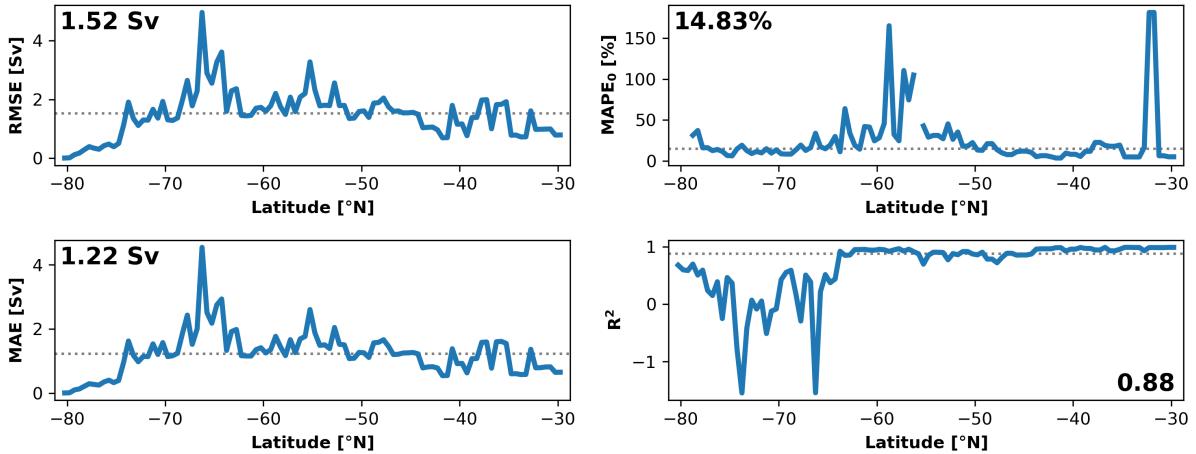


Figure 15: ElasticNet performance for independently trained models (one for each latitude) for all latitudes in the Southern Ocean with ECCO outputs (80.5°S to 30°S). Median test performance across all latitudes is indicated by the dotted line and the bold inset.  $\text{MAPE}_0$  values greater than 500% are removed from the plot (and median calculations) to facilitate interpretation.

### 3.5.2 Learning Latitude-Invariant Features

*Sharan Maiya*

We consider a new approach of training a single model over the whole Southern Ocean: a much larger dataset. Models are trained on the first 60% of all time series, validated on the next 20% and tested on the last 20%.

| Method     | RMSE $\downarrow$ | MAPE $_0\downarrow$ |
|------------|-------------------|---------------------|
| Linear Map | 2.44              | 27.39%              |
| MLP        | <b>2.28</b>       | <b>24.68%</b>       |
| RNN        | 4.52              | 33.70%              |
| GRU        | 5.10              | 62.50%              |
| LSTM       | 5.49              | 42.40%              |

Table 9: Median (over latitudes) test performance for models trained over the whole Southern Ocean (30°S to 80.5°S).

Table 9 presents performance metrics for five different architectures. Sequence models (RNN, GRU [7], LSTM [23]) are investigated in order to capture any temporal information which may lead to better generalisation - with training only feasible due to the larger dataset. However, mirroring the results in Section 3.3 for CNNs, these models perform poorly relative to simpler architectures.

In contrast, MLP and linear regression perform much better. We find the latter result particularly surprising; our linear map (trained using gradient-descent instead of least-squares) has a median RMSE of 2.44 across all latitudes - an RNN for comparison has a median RMSE of 4.52. This performance and earlier results constitute a body of empirical evidence that the dependencies between these satellite-observable variables and MOC strength can be approximated as linear.

Our top performing model is an MLP, striking a good balance between robustness (avoiding the overfitting issues of sequence models) and complexity (over linear regression). Several modelling adjustments are required to obtain the performance in Table 9. These are listed in Appendix A.4 to facilitate replication.

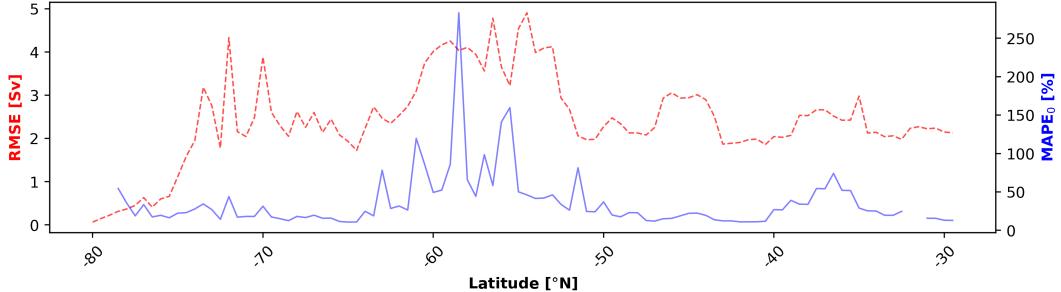


Figure 16: RMSE and MAPE $_0$  at all latitudes in the Southern Ocean for our trained MLP.

Figure 16 shows that our trained MLP performs worst around the latitude of 58°S, mirroring findings in Figures 14 and 15. This corresponds to the Drake Passage, which acts as a bottleneck for the ACC. Specifically, this is where the Sub-Antarctic Front and Polar Front meet [21, 86], and we believe the interactions here cause the abyssal MOC strength to be harder to predict.

These results clarify that it is indeed possible to learn spatially invariant predictive features of MOC strength from satellite-observable variables. We now consider the application of our approach to another dataset of higher resolution, in order to improve the reliability of our results.

### 3.6 Experiment 3: Using High-resolution Simulation Data (ACCESS)

*Tom Cowperthwaite*

ACCESS-OM2-01 is a state-of-the-art coupled climate model that accurately models oceanic and cryospheric processes (such as AABW formation) at high resolution [29]. How predictive model performance changes with this higher-resolution data remains an open question.

Here, we train models using satellite-observable variables obtained from historical runs of ACCESS (see Section 2.1). We repeat the regularised linear regression baseline and then train MLP models correspondingly to those trained in Section 3.2, only with a range of different network depths and widths (maximum width/depth = 5 neurons/layers). ZWS values used here have been obtained from ECCO. Further work could obtain the ZWS forcing directly from the JRA-55 reanalysis product that is used to drive ACCESS simulations [29, 82].

Table 10 shows results from fitting a 5-fold cross-validated ElasticNet regression on different input combinations (seasonality and trend retained) taken from 60°S [54]. These results are directly comparable to those presented using ECCO data in Table 2, and show that using high-spatial-resolution data generally improves predictive skill, with the best model trained on ACCESS data performing 9.42% (MAPE<sub>0</sub>) better than the most successful ECCO-trained model. We obtain even better results for predictions at 30°S, which is expected due to the apparent inherent predictive ease at that latitude, due to a dominant seasonal component (see Section 3.3.2) — full results are shown in Table 17. The rich resolution of eddy currents provided by ACCESS may be the cause of this increased performance, as eddies have been shown to be crucial driver of ocean circulation [16, 69].

| Inputs                                  | MAE↓        | RMSE↓       | Correlation↑ | R <sup>2</sup> ↑ | MAPE <sub>0</sub> ↓ | MAPE <sub>0.5</sub> ↓ |
|---|-------------|-------------|--------------|------------------|---------------------|-----------------------|
| all inputs, zonal avg.                  | <b>1.77</b> | <b>2.17</b> | <b>0.95</b>  | <b>0.88</b>      | 44.91%              | <b>19.47%</b>         |
| OBP, full zonal                         | 2.31        | 2.80        | 0.91         | 0.80             | 22.92%              | 26.95%                |
| OBP, full zonal + ZWS, zonal avg.       | <b>1.61</b> | <b>1.93</b> | <b>0.96</b>  | <b>0.91</b>      | <b>17.51%</b>       | <b>18.16%</b>         |
| OBP + SSH, full zonal + ZWS, zonal avg. | 2.25        | 2.73        | 0.92         | 0.84             | 37.88%              | 24.43%                |

Table 10: Performance metrics for regularised linear regression (ElasticNet) at 60°S.

With seasonality and trend retained, MLP models experience mode collapse, similarly to results in Section 3.2.3. In the deseasonalised, detrended case, the MLP models exhibit predictive skill, shown in Table 11. Both RMSE and MAPE<sub>0</sub> scores show that MLPs are unable to recover the performance of linear regression, despite inhabiting a larger hypothesis class. If trained on a larger dataset, we hypothesise that the MLP may rediscover the linear regression solution and improve on it.

| Inputs                                  | MAE↓        | RMSE↓       | Correlation↑ | R <sup>2</sup> ↑ | MAPE <sub>0</sub> ↓ | MAPE <sub>0.5</sub> ↓ |
|---|-------------|-------------|--------------|------------------|---------------------|-----------------------|
| OBP, full zonal                         | <b>2.44</b> | <b>2.95</b> | <b>0.86</b>  | <b>0.74</b>      | 185.69%             | 74.76%                |
| SSH, full zonal                         | 3.04        | 3.77        | 0.80         | 0.63             | 197.60%             | 98.51%                |
| OBP, full zonal + ZWS, zonal avg.       | <b>2.82</b> | <b>3.55</b> | <b>0.86</b>  | <b>0.73</b>      | 238.32%             | <b>70.64%</b>         |
| OBP + SSH, full zonal + ZWS, zonal avg. | 2.90        | 3.64        | <b>0.89</b>  | <b>0.77</b>      | <b>107.34%</b>      | 71.32%                |

Table 11: Model performance of MLP models at 60°S. In each case, the best performing model has one hidden layer with five units.

## 4 Conclusion

*Tom Cowperthwaite*

In this study, we advance the state-of-the-art in estimating MOC strength using only satellite-observable variables in four primary ways, as well as provide a rigorous baseline using linear regression. We find that the relationship between satellite-observable variables and MOC strength is approximately linear across latitudes (Sections 3.1 and 3.5.2).

In Section 3.3 we use CNNs to exploit the spatiotemporal properties of the MOC and ocean dynamics more generally. While the models offer some predictive skill, they perform comparably to linear regres-

sion, supporting the conclusion that the relationship between satellite-observed variables is approximately linear.

Next, in Section 3.4, we find that using non-local empirical data alone (RAPID MOC strength measurements) has some predictive power when applying linear regression, particularly when a time-delay is introduced. A time lag of  $>4$  years improves  $\text{MAPE}_0$  scores by  $\sim 7\%$ , compared to cotemporal RAPID data ( $\text{MAPE}_0 = 31.35\%$ ). This provides indication of a possible bi-polar seesaw, but further work using longer time frames is required to obtain robust evidence.

In Section 3.5, we train a suite of models on ECCO data from  $60^\circ\text{S}$ , and test them on the full range of latitudes in the Southern Ocean, to assess the generality of our models. Here we find that our models perform comparably across the Southern Ocean, with  $60^\circ\text{S}$  proving to be a particularly difficult latitude to predict. This finding supports the notion that our models are learning general physical relationships between satellite-observable variables and circulation strength that are applicable to other locations. Further work could include other covariates such as bathymetric depth, bottom form stress, or latitude in order to contextualise predictions.

Finally, in Section 3.6, we assess the effect of using high-resolution data as inputs to our models. Here, we find that like-for-like performance is generally improved when moving from  $1^\circ$  to  $0.1^\circ$  resolution, likely due to improved representation of eddies and other small-scale processes. Future work could investigate the effects of including meltwater runoff in projected climate scenarios as a covariate, potentially paving the way for more rigorous studies into possible ocean tipping points.

## 5 Code and Data Availability

*Emilio Luz-Ricca*

All code used in our analyses is hosted on the following public GitHub repository (v2.0 release): <https://github.com/ai4er-cdt/oceans-gtc/tree/v2.0>. An archived version of the repository is also available on Zenodo: <https://zenodo.org/doi/10.5281/zenodo.10817670> [39]. Raw ECCO outputs can be downloaded from NASA PO.DAAC: <https://podaac.jpl.nasa.gov/>; all entries used are detailed in the Appendix (Table 18). Extracted ECCO outputs (i.e., satellite-observable variables) and processed data products (i.e., streamfunctions and MOC strength time series) used in our analyses are archived on Zenodo: <https://zenodo.org/doi/10.5281/zenodo.10832705> [38]. Raw ACCESS outputs were shared with us directly by colleagues at the National Oceanography Centre (NOC); this data cannot be shared openly, but is available upon request. RAPID data can be downloaded from the NOC British Oceanographic Data Centre: <https://doi.org/10.5285/04c79ece-3186-349a-e063-6c86abc0158c>.

## References

- [1] Adcroft, A., Hill, C., Campin, J.-M., Marshall, J., & Heimbach, P. (2004). Overview of the formulation and numerics of the mit gcm. *Proceedings of the ECMWF seminar series on Numerical Methods, Recent developments in numerical methods for atmosphere and ocean modelling*, 139–149.
- [2] Bailey, S. T., Jones, C. S., Abernathey, R. P., Gordon, A. L., & Yuan, X. (2023). Water mass transformation variability in the Weddell Sea in ocean reanalyses. *Ocean Science*, 19(2), 381–402.
- [3] Ballarotta, M., Drijfhout, S., Kuhlbrodt, T., & Döös, K. (2013). The residual circulation of the southern ocean: Which spatio-temporal scales are needed? *Ocean Modelling*, 64, 46–55.
- [4] Bourassa, M. A., Meissner, T., Cerovecki, I., Chang, P. S., Dong, X., De Chiara, G., Donlon, C., Dukhovskoy, D. S., Elya, J., Fore, A., Fewings, M. R., Foster, R. C., Gille, S. T., Haus, B. K., Hristova-Veleva, S., Holbach, H. M., Jelenak, Z., Knaff, J. A., Kranz, S. A., ... Wentz, F. (2019). Remotely Sensed Winds and Wind Stresses for Marine Forecasting and Ocean Modeling. *Frontiers in Marine Science*, 6.
- [5] Chapron, B., Collard, F., & Ardhuin, F. (2005). Direct measurements of ocean surface velocity from space: Interpretation and validation. *Journal of Geophysical Research: Oceans*, 110(7).
- [6] Chen, W., Zhang, T., & Guan, L. (2018). Radiation Transfer in the Ocean and Ocean Color. In *Volume 8: Oceans* (pp. 43–78).
- [7] Cho, K., van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning phrase representations using rnn encoder-decoder for statistical machine translation. *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing*, 43–78.
- [8] Chylek, P., Folland, C. K., Lesins, G., & Dubey, M. K. (2010). Twentieth century bipolar seesaw of the Arctic and Antarctic surface air temperatures. *Geophysical Research Letters*, 37(8).
- [9] Davtian, N., & Bard, E. (2023). A new view on abrupt climate changes and the bipolar seesaw based on paleotemperatures from Iberian Margin sediments. *Proceedings of the National Academy of Sciences*, 120(12), e2209558120.
- [10] Dohan, K. (2017). Ocean surface currents from satellite data. *Journal of Geophysical Research: Oceans*, 122(4), 2647–2651.
- [11] Du, Y., Dong, X., Jiang, X., Zhang, Y., Zhu, D., Sun, Q., Wang, Z., Niu, X., Chen, W., Zhu, C., Jing, Z., Tang, S., Li, Y., Chen, J., Chu, X., Xu, C., Wang, T., He, Y., Han, B., ... Peng, S. (2021). Ocean surface current multiscale observation mission (OSCOM): Simultaneous measurement of ocean surface current, vector wind, and temperature. *Progress in Oceanography*, 193, 102531.
- [12] ECCO Consortium, Fukumori, I., Wang, O., Fenty, I., Forget, G., Heimbach, P., & Ponte, R. M. (2024). ECCO central estimate (version 4 release 4). Retrieved March 18, 2024, from <https://podaac.jpl.nasa.gov/ECCO>
- [13] Foresee, F. D., & Hagan, M. T. (1997). Gauss-newton approximation to bayesian learning. *Proceedings of International Conference on Neural Networks (ICNN'97)*, 3, 1930–1935.
- [14] Forget, G., Campin, J.-M., Heimbach, P., Hill, C., Ponte, R., & Wunsch, C. (2015). Ecco version 4: An integrated framework for non-linear inverse modeling and global ocean state estimation. *Geoscientific Model Development*, 8(10), 3071–3104.
- [15] Frajka-Williams, E., Ansorge, I. J., Baehr, J., Bryden, H. L., Chidichimo, M. P., Cunningham, S. A., Danabasoglu, G., Dong, S., Donohue, K. A., Elipot, S., Heimbach, P., Holliday, N. P., Hummels, R., Jackson, L. C., Karstensen, J., Lankhorst, M., Le Bras, I. A., Lozier, M. S., McDonagh, E. L., ... Wilson, C. (2019). Atlantic meridional overturning circulation: Observed transport and variability. *Frontiers in Marine Science*, 6.
- [16] Frey, D. I., Morozov, E. G., & Smirnova, D. A. (2023). Sea level anomalies affect the ocean circulation at abyssal depths [Publisher: Nature Publishing Group]. *Scientific Reports*, 13(1), 20829. <https://doi.org/10.1038/s41598-023-48074-9>
- [17] Gent, P. R., & McWilliams, J. C. (1990). Isopycnal mixing in ocean circulation models. *Journal of Physical Oceanography*, 20(1), 150–155.
- [18] Gill, A. (1982). *Atmosphere-Ocean Dynamics* (1st ed.). Elsevier.
- [19] Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.
- [20] Griffies, S. M. (2012). *Elements of the modular ocean model (mom)* (tech. rep. No. 7). NOAA Geophysical Fluid Dynamics Laboratory Ocean Group.
- [21] Griffiths, H. (2010). *Zoogeographic and richness patterns in southern ocean benthos* [Doctoral dissertation, Open University].

- [22] Gruber, N., Bakker, D. C. E., DeVries, T., Gregor, L., Hauck, J., Landschützer, P., McKinley, G. A., & Müller, J. D. (2023). Trends and variability in the ocean carbon sink. *Nature Reviews Earth and Environment*, 4(2), 119–134.
- [23] Hochreiter, S., & Schmidhuber, J. (1997). Long Short-Term Memory. *Neural Computation*, 9(8), 1735–1780.
- [24] Hodson, T. O. (2022). Root-mean-square error (RMSE) or mean absolute error (MAE): When to use them or not. *Geoscientific Model Development*, 15(14), 5481–5487.
- [25] Jeffree, J. (2022). *A first estimate of global Antarctic Bottom Water transport variability from GRACE satellite data* [BPhil (Hons)]. The Australian National University.
- [26] Kanzow, T., Cunningham, S. A., Johns, W. E., Hirschi, J. J.-M., Marotzke, J., Baringer, M. O., Meinen, C. S., Chidichimo, M. P., Atkinson, C., Beal, L. M., Bryden, H. L., & Collins, J. (2010). Seasonal variability of the atlantic meridional overturning circulation at 26.5°n. *Journal of Climate*: 23(21), 5678–5698.
- [27] Karunasingha, D. S. K. (2022). Root mean square error or mean absolute error? Use their ratio as well. *Information Sciences*, 585, 609–629.
- [28] Kim, S., & Kim, H. (2016). A new metric of absolute percentage error for intermittent demand forecasts. *International Journal of Forecasting*, 32(3), 669–679.
- [29] Kiss, A. E., Hogg, A. M., Hannah, N., Boeira Dias, F., Brassington, G. B., Chamberlain, M. A., Chapman, C., Dobrohotoff, P., Domingues, C. M., Duran, E. R., et al. (2020a). Access-om2 v1. 0: A global ocean–sea ice model at three resolutions. *Geoscientific Model Development*, 13(2), 401–442.
- [30] Kiss, A. E., Hogg, A. M., Hannah, N., Boeira Dias, F., Brassington, G. B., Chamberlain, M. A., Chapman, C., Dobrohotoff, P., Domingues, C. M., Duran, E. R., England, M. H., Fiedler, R., Griffies, S. M., Heerdegen, A., Heil, P., Holmes, R. M., Klocker, A., Marsland, S. J., Morrison, A. K., ... Zhang, X. (2020b). ACCESS-OM2 v1.0: A global ocean–sea ice model at three resolutions. *Geoscientific Model Development*, 13(2), 401–442.
- [31] LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11), 2278–2324.
- [32] Li, G., Cheng, L., Zhu, J., Trenberth, K. E., Mann, M. E., & Abraham, J. P. (2020). Increasing ocean stratification over the past half-century. *Nature Climate Change*, 10(12), 1116–1123.
- [33] Li, Q., England, M. H., Hogg, A. M., Rintoul, S. R., & Morrison, A. K. (2023). Abyssal ocean overturning slowdown and warming driven by antarctic meltwater. *Nature*, 615(7954), 841–847.
- [34] Li, X., Cai, W., Meehl, G. A., Chen, D., Yuan, X., Raphael, M., Holland, D. M., Ding, Q., Fogt, R. L., Markle, B. R., Wang, G., Bromwich, D. H., Turner, J., Xie, S.-P., Steig, E. J., Gille, S. T., Xiao, C., Wu, B., Lazzara, M. A., ... Song, C. (2021). Tropical teleconnection impacts on Antarctic climate changes. *Nature Reviews Earth & Environment*, 2(10), 680–698.
- [35] Liu, M., & Tanhua, T. (2021). Water masses in the Atlantic Ocean: Characteristics and distributions. *Ocean Science*, 17(2), 463–486.
- [36] Liu, W., Xie, S.-P., Liu, Z., & Zhu, J. (2017). Overlooked possibility of a collapsed Atlantic Meridional Overturning Circulation in warming climate. *Science Advances*, 3(1), e1601666.
- [37] Loshchilov, I., & Hutter, F. (2017). Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*.
- [38] Luz-Ricca, E., Baranduin, N., Cowperthwaite, T., Maiya, S., & Van Driessche, A. (2024a). *Ai4er gtc: Ecco streamfunction, moc strength, and surface variables* (tech. rep.). University of Cambridge. <https://doi.org/10.5281/zenodo.10832705>
- [39] Luz-Ricca, E., Baranduin, N., Cowperthwaite, T., Maiya, S., & Van Driessche, A. (2024b, March). *Oceans-gtc: Ai4er gtc github repository* (Version 2.0). <https://doi.org/10.5281/zenodo.10817670>
- [40] MacKay, D. J. C. (1992). Bayesian Interpolation. *Neural Computation*, 4(3), 415–447.
- [41] Majumder, S., Schmid, C., & Halliwell, G. (2016). An observations and model-based analysis of meridional transports in the South Atlantic. *Journal of Geophysical Research: Oceans*, 121(8), 5622–5638.
- [42] McCarthy, G. D., Brown, P. J., Flagg, C. N., Goni, G., Houpert, L., Hughes, C. W., Hummels, R., Inall, M., Jochumsen, K., Larsen, K. M. H., Lherminier, P., Meinen, C. S., Moat, B. I., Rayner, D., Rhein, M., Roessler, A., Schmid, C., & Smeed, D. A. (2020). Sustainable Observations of the AMOC: Methodology and Technology. *Reviews of Geophysics*, 58(1), e2019RG000654.
- [43] Meredith, M. P., Schofield, O., Newman, L., Urban, E., & Sparrow, M. (2013). The vision for a Southern Ocean Observing System. *Current Opinion in Environmental Sustainability*, 5(3), 306–313.
- [44] Moat, B. I., Smeed, D., Rayner, D., Johns, W. E., Smith, R. H., Volkov, D. L., Baringer, M. O., & Collins, J. (2023, September). Atlantic meridional overturning circulation observed by the RAPID-

- MOCHA-WBTS (RAPID-Meridional Overturning Circulation and Heatflux Array-Western Boundary Time Series) array at 26N from 2004 to 2022 (v2022.1). Retrieved March 18, 2024, from <https://nora.nerc.ac.uk/id/eprint/535847/>
- [45] Montaño Moreno, J. J., Palmer Pol, A., & Sesé Abad, A. (2013). Using the R-MAPE index as a resistant measure of forecast accuracy. *Psicothema*, (25), 500–506.
- [46] Morley, S. K., Brito, T. V., & Welling, D. T. (2018). Measures of Model Performance Based On the Log Accuracy Ratio. *Space Weather*, 16(1), 69–88.
- [47] Msadek, R., Johns, W. E., Yeager, S. G., Danabasoglu, G., Delworth, T. L., & Rosati, A. (2013). The atlantic meridional heat transport at 26.5°n and its relationship with the moc in the rapid array and the gfdl and ncar coupled models. *Journal of Climate*, 26(12), 4335–4356.
- [48] Muir, L. C., & Fedorov, A. V. (2015). How the AMOC affects ocean temperatures on decadal to centennial timescales: The North Atlantic versus an interhemispheric seesaw. *Climate Dynamics*, 45(1), 151–160.
- [49] Muller-Karger, F. E., Smith, J. P., Werner, S., Chen, R., Roffer, M., Liu, Y., Muhling, B., Lindo-Atchati, D., Lamkin, J., Cerdeira-Estrada, S., & Enfield, D. B. (2015). Natural variability of surface oceanographic conditions in the offshore Gulf of Mexico. *Progress in Oceanography*, 134, 54–76.
- [50] Nadeau, L.-P., & Jansen, M. F. (2020). Overturning Circulation Pathways in a Two-Basin Ocean Model. *Journal of Physical Oceanography*, 50(8), 2105–2122.
- [51] Nair, V., & Hinton, G. E. (2010). Rectified linear units improve restricted boltzmann machines. *Proceedings of the 27th international conference on machine learning (ICML-10)*, 807–814.
- [52] Pao, H. T. (2009). Forecasting energy consumption in Taiwan using hybrid nonlinear models. *Energy*, 34(10), 1438–1446.
- [53] Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Kopf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., ... Chintala, S. (2019). Pytorch: An imperative style, high-performance deep learning library. In *Advances in neural information processing systems 32* (pp. 8024–8035). Curran Associates, Inc.
- [54] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.
- [55] Pedro, J. B., Jochum, M., Buizert, C., He, F., Barker, S., & Rasmussen, S. O. (2018). Beyond the bipolar seesaw: Toward a process understanding of interhemispheric coupling. *Quaternary Science Reviews*, 192, 27–46.
- [56] Pelichero, V., Sallée, J.-B., Schmidtko, S., Roquet, F., & Charrassin, J.-B. (2017). The ocean mixed layer under southern ocean sea-ice: Seasonal cycle and forcing. *Journal of Geophysical Research - Oceans*, 122(2), 1608–1633.
- [57] Previdi, M., Smith, K. L., & Polvani, L. M. (2021). Arctic amplification of climate change: A review of underlying mechanisms [Publisher: IOP Publishing]. *Environmental Research Letters*, 16(9), 093003.
- [58] Rahmstorf, S. (2003). Thermohaline circulation: The current climate. *Nature*, 421(6924), 699–699.
- [59] Rivaro, P., Massolo, S., Bergamasco, A., Castagno, P., & Budillon, G. (2010). Chemical evidence of the changes of the Antarctic Bottom Water ventilation in the western Ross Sea between 1997 and 2003. *Deep Sea Research Part I: Oceanographic Research Papers*, 57(5), 639–652.
- [60] Robinson, I. S. (2010). The methods of satellite oceanography. In I. S. Robinson (Ed.), *Discovering the Ocean from Space: The unique applications of satellite oceanography* (pp. 7–67). Springer.
- [61] Roemmich, D., & Wunsch, C. (1985). Two transatlantic sections: Meridional circulation and heat flux in the subtropical north atlantic ocean. *Deep Sea Research Part A. Oceanographic Research Papers*, 32(6), 619–664. [https://doi.org/https://doi.org/10.1016/0198-0149\(85\)90070-6](https://doi.org/https://doi.org/10.1016/0198-0149(85)90070-6)
- [62] Rousselet, L., & Cessi, P. (2022). Diabatic Transformations along the Global Routes of the Mid-depth Meridional Overturning Circulation. *Journal of Physical Oceanography*, 52(12), 3159–3177.
- [63] Rousselet, L., Cessi, P., & Forget, G. (2021a). Coupling of the mid-depth and abyssal components of the global overturning circulation according to a state estimate. *Science Advances*, 7(21), eabf5478.
- [64] Rousselet, L., Cessi, P., & Forget, G. (2021b). Coupling of the mid-depth and abyssal components of the global overturning circulation according to a state estimate. *Science Advances*, 7(21), eabf5478.
- [65] Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, 323(6088), 533–536.

- [66] Sarmiento, J. L., & Gruber, N. (2006, June). *Ocean Biogeochemical Dynamics*. Princeton University Press.
- [67] Seabold, S., & Perktold, J. (2010). Statsmodels: Econometric and statistical modeling with python. *9th Python in Science Conference*.
- [68] Shcherbakov, M., Brebels, A., Shcherbakova, N., Tyukov, A., Janovsky, T., & Kamaev, V. (2013). A survey of forecast error measures. *World Applied Sciences Journal*, 24, 171–176.
- [69] Sheen, K. L., Naveira Garabato, A. C., Bearley, J. A., Meredith, M. P., Polzin, K. L., Smeed, D. A., Forryan, A., King, B. A., Sallée, J.-B., St. Laurent, L., Thurnherr, A. M., Toole, J. M., Waterman, S. N., & Watson, A. J. (2014). Eddy-induced variability in Southern Ocean abyssal mixing on climatic timescales [Publisher: Nature Publishing Group]. *Nature Geoscience*, 7(8), 577–582. <https://doi.org/10.1038/ngeo2200>
- [70] Silvano, A., Purkey, S., Gordon, A. L., Castagno, P., Stewart, A. L., Rintoul, S. R., Foppert, A., Gunn, K. L., Herraiz-Borreguero, L., Aoki, S., Nakayama, Y., Naveira Garabato, A. C., Spingys, C., Akhoudas, C. H., Sallée, J.-B., de Lavergne, C., Abrahamsen, E. P., Meijers, A. J. S., Meredith, M. P., ... Lee, W. S. (2023). Observing antarctic bottom water in the southern ocean. *Frontiers in Marine Science*, 10. <https://doi.org/10.3389/fmars.2023.1221701>
- [71] Solodoch, A., Stewart, A. L., McC. Hogg, A., & Manucharyan, G. E. (2023). Machine learning-derived inference of the meridional overturning circulation from satellite-observable variables in an ocean state estimate. *Journal of Advances in Modeling Earth Systems*, 15(4), e2022MS003370.
- [72] Solodoch, A., & Stewart, A. L. (2022). Machine learning reconstruction of the meridional overturning circulation in the eccov4r3 ocean state estimate. <https://doi.org/10.5281/zenodo.701624>
- [73] Stepanov, V. N., Haines, K., & Smith, G. C. (2012). Assimilation of RAPID array observations into an ocean model. *Quarterly Journal of the Royal Meteorological Society*, 138(669), 2105–2117.
- [74] Stewart, A. L., Chi, X., Solodoch, A., & Hogg, A. M. (2021). High-frequency Fluctuations in Antarctic Bottom Water Transport Driven by Southern Ocean Winds. *Geophysical Research Letters*, 48(17).
- [75] Stewart, A. L., & Hogg, A. M. (2017). Reshaping the Antarctic Circumpolar Current via Antarctic Bottom Water Export. *Journal of Physical Oceanography*, 47(10), 2577–2601.
- [76] Tagklis, F., Bracco, A., Ito, T., & Castelao, R. M. (2020). Submesoscale modulation of deep water formation in the Labrador Sea. *Scientific Reports*, 10, 17489.
- [77] Tan, S., & Thurnherr, A. M. (2023). On the Global Decrease in the Deep and Abyssal Density Stratification Along the Spreading Pathways of Antarctic Bottom Water Since the 1990s. *Geophysical Research Letters*, 50(11), e2022GL102422.
- [78] Tapley, B. D., Bettadpur, S., Watkins, M., & Reigber, C. (2004). The gravity recovery and climate experiment: Mission overview and early results. *Geophysical Research Letters*, 31(9).
- [79] Thomas, A. C., Pershing, A. J., Friedland, K. D., Nye, J. A., Mills, K. E., Alexander, M. A., Record, N. R., Weatherbee, R., & Henderson, M. E. (2017). Seasonal trends and phenology shifts in sea surface temperature on the North American northeastern continental shelf. *Elementa: Science of the Anthropocene*, 5, 48.
- [80] Timmermans, M.-L., & Marshall, J. (2020). Understanding Arctic Ocean Circulation: A Review of Ocean Dynamics in a Changing Climate. *Journal of Geophysical Research: Oceans*, 125(4), e2018JC014378.
- [81] Tofallis, C. (2015). A better measure of relative prediction accuracy for model selection and model estimation. *Journal of the Operational Research Society*, 66(8), 1352–1362.
- [82] Tsujino, H., Urakawa, S., Nakano, H., Small, R. J., Kim, W. M., Yeager, S. G., Danabasoglu, G., Suzuki, T., Bamber, J. L., Bentsen, M., Böning, C. W., Bozec, A., Chassignet, E. P., Curchitser, E., Boeira Dias, F., Durack, P. J., Griffies, S. M., Harada, Y., Ilicak, M., ... Yamazaki, D. (2018). JRA-55 based surface dataset for driving ocean-sea-ice models (JRA55-do). *Ocean Modelling*, 130, 79–139.
- [83] Verbesselt, J., Hyndman, R., Zeileis, A., & Culvenor, D. (2010). Phenological change detection while accounting for abrupt and gradual trends in satellite image time series. *Remote Sensing of Environment*, 114(12), 2970–2980.
- [84] Wang, O., Fukumori, I., & Fenty, I. (2017). An overview of ecco version 4 release 3's ftp site <ftp://ecco.jpl.nasa.gov/version4/release3>.
- [85] Weijer, W., Cheng, W., Drijfhout, S. S., Fedorov, A. V., Hu, A., Jackson, L. C., Liu, W., McDonagh, E. L., Mecking, J. V., & Zhang, J. (2019). Stability of the Atlantic Meridional Overturning Circulation: A Review and Synthesis. *Journal of Geophysical Research: Oceans*, 124(8), 5336–5375.

- [86] Wu, S., Kuhn, G., Diekmann, B., Lembke-Jene, L., Tiedemann, R., Zheng, X., Ehrhardt, S., Arz, H., & Lamy, F. (2019). Surface sediment characteristics related to provenance and ocean circulation in the drake passage sector of the southern ocean. *Deep Sea Research Part I Oceanographic Research Papers*, 154, 103135.
- [87] Zhang, G. P., & Qi, M. (2005). Neural network forecasting for seasonal and trend time series. *European Journal of Operational Research*, 160(2), 501–514.
- [88] Zhao, K., Wulder, M. A., Hu, T., Bright, R., Wu, Q., Qin, H., Li, Y., Toman, E., Mallick, B., Zhang, X., & Brown, M. (2019). Detecting change-point, trend, and seasonality in satellite time series data to track abrupt changes and nonlinear dynamics: A Bayesian ensemble algorithm. *Remote Sensing of Environment*, 232, 111181.
- [89] Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 67(2), 301–320.

## A Appendices

### A.1 Depth-to-Density Conversion

The integrated velocity calculations output a result in depth-space (physical distance from the sea surface) but as ocean layers are defined by their properties rather than the physical depth measurements, we convert this depth space to density space. Given the salinity  $S$  and potential temperature  $\Theta$ , which are both input parameters available in ECCO, we calculate the conservative temperature  $CT$  using a thermodynamic equation of state for seawater [15]. Subsequently, the potential density anomaly  $\sigma_2$  at a reference pressure corresponding to a depth of 2 km depends on the salinity  $S$  and conservative temperature  $CT$ .

$$CT = f(S, \Theta)$$

$$\sigma_2 = g(S, CT)$$

### A.2 Evaluation Metrics: Full Definitions

If  $\Psi_p(t)$  and  $\Psi_m(t)$  are the predicted and actual MOC strength values at time  $t$ , respectively, with  $N_t$  test time steps in total, then our evaluation metrics are defined as:

$$\text{MAE} = \frac{1}{N_t} \sum_{t=1}^{N_t} |\Psi_m(t) - \Psi_p(t)| \quad (3)$$

$$\text{RMSE} = \sqrt{\frac{1}{N_t} \sum_{t=1}^{N_t} (\Psi_m(t) - \Psi_p(t))^2} \quad (4)$$

$$\text{MAPE}_\tau = \frac{1}{N_t} \sum_{k \in K_\tau} 100 \cdot \left| \frac{\Psi_m(k) - \Psi_p(k)}{\Psi_m(k) + \epsilon} \right| \quad (5)$$

$$r = \frac{\sum_{t=1}^{N_t} (\Psi_m(t) - \bar{\Psi}_m)(\Psi_p(t) - \bar{\Psi}_p)}{\sqrt{\sum_{t=1}^{N_t} (\Psi_m(t) - \bar{\Psi}_m)^2} \sqrt{\sum_{t=1}^{N_t} (\Psi_p(t) - \bar{\Psi}_p)^2}} \quad (6)$$

where  $\bar{\Psi}_m$  and  $\bar{\Psi}_p$  are time means for the actual and predicted MOC strength values, respectively. In the calculation of  $\text{MAPE}_\tau$ ,  $\epsilon$  is a small number to avoid dividing by zero and  $\tau$  controls the cutoff threshold for values considered in the calculation. More precisely,  $K_\tau$  is defined as:

$$K_\tau = \{t \in \{1, \dots, N_t\} : |\Psi_m(t)| > \tau\} \quad (7)$$

This helps reduce the outsized effect of very small values, which can significantly inflate the MAPE score; we report results for  $\tau \in \{0, 0.5\}$ . For linear regression models, we also use the coefficient of determination ( $R^2$ ):

$$R^2 = 1 - \frac{\sum_{t=1}^{N_t} (\Psi_m(t) - \Psi_p(t))^2}{\sum_{t=1}^{N_t} (\Psi_m(t) - \bar{\Psi}_m)^2} \quad (8)$$

Correlation measures performance by checking linear agreement between actual and predicted MOC time series. RMSE and MAE can be interpreted in the native units (i.e., Sv); RMSE penalizes larger errors more strongly than MAE. However, care must be taken when interpreting RMSE and MAE for models trained at different latitudes, since the distribution of abyssal MOC strength time series shows clear variations in both mean and variance across the four latitudes of interest (Figure 17). Therefore we use  $\text{MAPE}_\tau$ , a measure of relative error that facilitates comparison of performance across latitudes. Finally,  $R^2$  can be interpreted as the proportion of variance explained by a linear regression model; in general, it is a comparison of model predictive performance over predicting the time series mean.

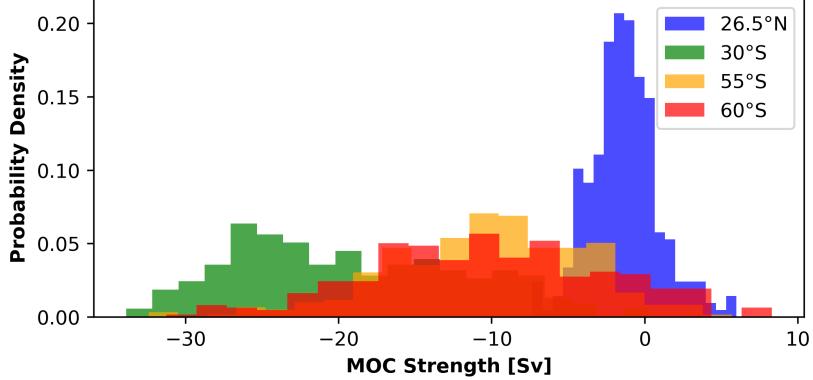


Figure 17: Distribution of abyssal MOC strength time series for the four latitudes of interest. We find significant differences in both mean and variance, which may lead to misinterpretation of scale-aware performance metrics (i.e., RMSE and MAE) when comparing between different latitudes.

### A.3 Additional Results

#### A.3.1 Baseline Modelling: Linear Regression

| Inputs                            | 26.5°N         | 30°S           | 55°S          | 60°S          |
|-----------------------------------|----------------|----------------|---------------|---------------|
| all inputs, zonal avg.            | 1245.90%       | <b>112.18%</b> | 174.55%       | 108.93%       |
| OBP + SSH + ZWS, full zonal       | <b>61.20%</b>  | 1384.80%       | <b>67.87%</b> | 103.68%       |
| OBP, full zonal                   | 149.58%        | 735.23%        | <u>59.01%</u> | <u>57.58%</u> |
| OBP, full zonal + ZWS, zonal avg. | <u>134.37%</u> | 620.53%        | <b>58.55%</b> | <b>53.39%</b> |

Table 12: Performance as measured by  $\text{MAPE}_0$  (lower is better) for ElasticNet at all latitudes of interest (as defined in Section 2.2). Model inputs and outputs are detrended and deseasonalized. Best results are in bold and second best are underlined.

| Inputs                                | 26.5°N        | 30°S         | 55°S          | 60°S          |
|---------------------------------------|---------------|--------------|---------------|---------------|
| all inputs, zonal avg.                | 392.54%       | 14.59%       | 56.22%        | 101.31%       |
| all inputs, 6-month lags, zonal avg.  | 443.54%       | 14.97%       | 45.94%        | 69.73%        |
| all inputs, 12-month lags, zonal avg. | 312.33%       | 16.84%       | 47.47%        | 74.96%        |
| OBP + SSH + ZWS, full zonal           | 94.43%        | <b>9.32%</b> | <u>23.09%</u> | 39.57%        |
| OBP, full zonal                       | <u>46.74%</u> | 11.27%       | <b>20.46%</b> | <u>35.88%</u> |
| OBP, full zonal + ZWS, zonal avg.     | <b>43.68%</b> | <u>9.96%</u> | 26.81%        | <b>34.68%</b> |

Table 13: Performance as measured by  $\text{MAPE}_0$  (lower is better) for ElasticNet at all latitudes of interest (as defined in Section 2.2). Best results are in bold and second best are underlined.

#### A.3.2 Model Advancements: Trend and Seasonality

To assess how removing or including seasonality and trends impacted model performance, neural networks were ran using a test train split of 80% and 20% respectively. Hyper-parameters used in each model are included in 14.

| Latitude | dropout | pure layers | mix layers | features | channels | kernel |
|----------|---------|-------------|------------|----------|----------|--------|
| 26.5°N   | 0.5     | 5           | 1          | 5        | 3        | 3      |
| 30°S     | 0.5     | 5           | 1          | 5        | 5        | 5      |
| 55°S     | 0.5     | 5           | 1          | 5        | 3        | 3      |
| 60°S     | 0.5     | 5           | 1          | 5        | 5        | 5      |

Table 14: Hyper-parameters used for each latitude in the 1D-CNNs that were used to assess how removing or including seasonality and trend in the data impacted model performance.

### A.3.3 Model Advancements: 1D-CNNs

The overall best performing 1D-CNN network at 55°S (and 30° for the Atlantic basin) consists of 3 convolutional layers where each of these layers will produce 5 distinct feature maps. The kernel size is 3, dropout is 0.1 and other parameters such as learning rate and weight decay are as described in 3.2.

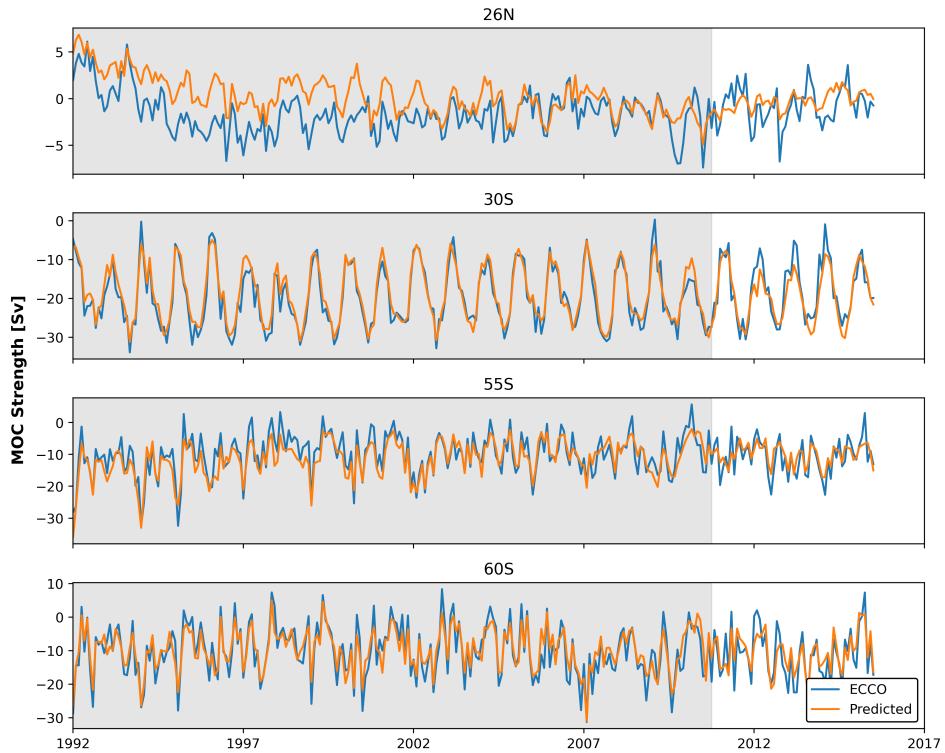


Figure 18: Predictions vs MOC strength for 1D-CNN models on each of the 4 latitudes. Training and validation data are plotted against a grey background.

### A.3.4 Incorporating Observational Data (RAPID)

| Inputs      | timelag                   | MAE $\downarrow$ | RMSE $\downarrow$ | MAPE $_{0.0}\downarrow$ | MAPE $_{0.5}\downarrow$ | Correlation $\uparrow$ |
|-------------|---------------------------|------------------|-------------------|-------------------------|-------------------------|------------------------|
| ZWS         | -                         | 0.67             | 0.85              | 16.26%                  | 16.26%                  | 0.77                   |
| ZWS + RAPID | -                         | 0.67             | 0.85              | 16.23%                  | 16.23%                  | 0.73                   |
| ZWS + RAPID | 12 months                 | 0.62             | 0.65              | 16.53%                  | 16.53%                  | <b>0.78</b>            |
| ZWS + RAPID | 12, 15, 18, 21, 24 months | <b>0.65</b>      | <b>0.80</b>       | <b>15.46%</b>           | <b>15.46%</b>           | 0.77                   |

Table 15: Model Performance for Linear Regression on the 30°S bottom cell in the Atlantic Ocean. The first row forms the baseline reference, comparing to the inclusion of RAPID data with various timelags (and multiple timelags combined).

| Inputs                        | timelag  | MAE $\downarrow$ | RMSE $\downarrow$ | MAPE $_0\downarrow$ | MAPE $_{0.5}\downarrow$ | Correlation $\uparrow$ |
|-------------------------------|----------|------------------|-------------------|---------------------|-------------------------|------------------------|
| OBP + ZWS, zonal avg.         | -        | 0.95             | 0.75              | 14.70%              | 14.70%                  | <b>0.77</b>            |
| OBP + ZWS + RAPID, zonal avg. | -        | 0.90             | 0.74              | 15.10%              | 15.10%                  | 0.73                   |
| OBP + ZWS + RAPID, zonal avg. | 6 months | <b>0.79</b>      | <b>0.61</b>       | <b>12.70%</b>       | <b>12.70%</b>           | 0.75                   |
| OBP + ZWS + RAPID, full zonal | 3 months | 1.23             | 0.99              | 18.5%               | 18.5%                   | 0.49                   |

Table 16: Model Performance for Neural Networks on the 30°S bottom cell in the Atlantic Ocean. The first row forms the baseline reference, comparing to the inclusion of RAPID data with various time lags.

### A.3.5 Linear Regression Results for 30°S (ACCESS)

| Inputs                                  | MAE $\downarrow$ | RMSE $\downarrow$ | Correlation $\uparrow$ | $R^2 \uparrow$ | MAPE $_0 \downarrow$ | MAPE $_{0.5} \downarrow$ |
|---|------------------|-------------------|------------------------|----------------|----------------------|--------------------------|
| OBP, full zonal                         | <u>0.68</u>      | <u>0.85</u>       | <u>0.82</u>            | <u>0.48</u>    | <u>18.42%</u>        | <u>16.56%</u>            |
| SSH, full zonal                         | 1.00             | 1.26              | 0.48                   | -1.80          | 22.06%               | 29.06%                   |
| OBP, full zonal + ZWS, zonal avg.       | <b>0.60</b>      | <b>0.76</b>       | <b>0.85</b>            | <b>0.57</b>    | <b>15.54%</b>        | <b>14.85%</b>            |
| OBP + SSH, full zonal + ZWS, zonal avg. | 0.72             | 0.86              | 0.80                   | 0.42           | 18.20%               | 18.03%                   |

Table 17: Performance metrics for regularised linear regression (ElasticNet) at 30°S, with different input variables. For each metric shown, the best performing combination of variables is shown in bold, with the second-best underlined.

## A.4 Experiment Details for Full Southern Ocean Methods

The model results in Table 9 are obtained after several adjustments to the original modelling setup. These are includes below, to facilitate future replication:

- A composite loss function combining MAE and MSE, with L1 and L2 regularisation is used.
- Full zonal information of SSH, SSS, SST, OBP, and ZWS is used as input; the removal of any immediately results in a performance drop, suggesting that at this dataset size models are able to fully leverage information from all signals.
- The addition of  $\mathcal{N}(0, 0.1)$  Gaussian noise improves generalisation.
- Increased model depth as opposed to width improves model performance - all trained neural networks have 10 hidden layers and widths between 16 and 32 neurons.
- Training our networks with dropout actively hurts model performance.

## A.5 Notes on ECCO Outputs

All ECCO data was downloaded from NASA PO.DAAC programatically through Python using the NASA Common Metadata Repository (CMR). See Table 18 for the relevant PO.DAAC entry names and corresponding DOIs.

| Data Type             | PO.DAAC Entry                            | DOI   |
|-----------------------|--|---|
| SSH                   | ECCO_L4 SSH_05DEG_MONTHLY_V4R4           | <a href="https://doi.org/10.5067/ECG5M-SSH44">https://doi.org/10.5067/ECG5M-SSH44</a> |
| SSS                   | ECCO_L4 TEMP_SALINITY_05DEG_MONTHLY_V4R4 | <a href="https://doi.org/10.5067/ECG5M-OTS44">https://doi.org/10.5067/ECG5M-OTS44</a> |
| SST                   | ECCO_L4 TEMP_SALINITY_05DEG_MONTHLY_V4R4 | <a href="https://doi.org/10.5067/ECG5M-OTS44">https://doi.org/10.5067/ECG5M-OTS44</a> |
| ZWS                   | ECCO_L4 STRESS_05DEG_MONTHLY_V4R4        | <a href="https://doi.org/10.5067/ECG5M-STR44">https://doi.org/10.5067/ECG5M-STR44</a> |
| OBP                   | ECCO_L4 OBP_05DEG_MONTHLY_V4R4           | <a href="https://doi.org/10.5067/ECG5M-OPB44">https://doi.org/10.5067/ECG5M-OPB44</a> |
| Monthly-Mean Velocity | ECCO_L4 OCEAN_VEL_05DEG_MONTHLY_V4R4     | <a href="https://doi.org/10.5067/ECG5M-OVE44">https://doi.org/10.5067/ECG5M-OVE44</a> |
| Bolus Velocity        | ECCO_L4 BOLUS_05DEG_MONTHLY_V4R4         | <a href="https://doi.org/10.5067/ECG5M-BOL44">https://doi.org/10.5067/ECG5M-BOL44</a> |
| Model Grid Geometry   | ECCO_L4 GEOMETRY_LLC0090GRID_V4R4        | <a href="https://doi.org/10.5067/ECL5A-GRD44">https://doi.org/10.5067/ECL5A-GRD44</a> |

Table 18: All ECCO outputs used in our analysis. All fields were extracted as monthly means from 1/1/1992 to 31/12/2017 on the provided 0.5° interpolated latitude-longitude grid. DOIs are included for completeness.