JAMES | Journal of Advances in Modeling Earth Systems

RESEARCH ARTICLE
10.1029/2024MS004915

Full-Depth Reconstruction of Long-Term Meridional Overturning Circulation Variability From Satellite-Measurable Quantities via Machine Learning

Huaiyu Wei[1], Kaushik Srinivasan[1], Andrew L. Stewart[1], Aviv Solodoch[2], Georgy E. Manucharyan[3], and Andrew McC. Hogg[4]

[1]Department of Atmospheric and Oceanic Sciences, University of California, Los Angeles, Los Angeles, CA, USA, [2]Institute of Earth Sciences, The Hebrew University of Jerusalem, Jerusalem, Israel, [3]School of Oceanography, University of Washington, Seattle, WA, USA, [4]Research School of Earth Sciences, Australian National University, Canberra, ACT, Australia

Key Points:

- We developed a non-local dual-branch neural network to reconstruct the long-term variability of the meridional overturning circulation (MOC)
- Ocean bottom pressure near western boundaries and along dense water export pathways are identified as the dominant input features
- Latitudinally non-local ocean bottom pressure information is crucial for equatorial Atlantic MOC reconstruction

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

H. Wei,
hywei@atmos.ucla.edu

Citation:

Wei, H., Srinivasan, K., Stewart, A. L., Solodoch, A., Manucharyan, G. E., & Hogg, A. M. (2025). Full-depth reconstruction of long-term meridional overturning circulation variability from satellite-measurable quantities via machine learning. *Journal of Advances in Modeling Earth Systems*, *17*, e2024MS004915. https://doi.org/10.1029/2024MS004915

Received 19 DEC 2024
Accepted 21 JUN 2025
Corrected 18 OCT 2025

This article was corrected on 18 OCT 2025. See the end of the full text for details.

Author Contributions:

Conceptualization: Andrew L. Stewart, Aviv Solodoch, Georgy E. Manucharyan, Andrew McC. Hogg
Data curation: Huaiyu Wei
Formal analysis: Huaiyu Wei
Funding acquisition: Andrew L. Stewart

Abstract The meridional overturning circulation (MOC) plays a crucial role in the global distribution of heat, carbon, and other climate-relevant tracers. Monitoring the evolution of MOC is essential for understanding climate variability, yet direct MOC observations are sparse and geographically limited. Although satellite measurements have shown potential for short-term monitoring of the MOC, it remains unclear whether MOC variability on decadal and longer timescales can be detected remotely. In this study, we leverage machine learning to reconstruct long-term MOC variability from satellite-measurable quantities, using climate simulations under pre-industrial conditions. We demonstrate that our proposed non-local dual-branch neural network (DBNN) effectively reconstructs both the strength and vertical structure of the Atlantic MOC (AMOC) and the Southern Ocean MOCs across sub-annual to multi-decadal timescales. Using a neural network interpretation technique, we identify ocean bottom pressure near the western boundary and along dense-water export pathways as the dominant input features for MOC reconstruction. This indicates that DBNN's predictions can be interpreted as an approximation of geostrophic balance. The DBNN also effectively reconstructs the AMOC in the equatorial region, where geostrophy breaks down. This success is attributed to the capability of DBNN in utilizing latitudinally non-local ocean bottom pressure information and the meridional coherence of AMOC variability. Additionally, the DBNN accurately reconstructs Southern Ocean MOCs using only sea surface height and zonal wind stress as inputs, thereby avoiding reliance on ocean bottom pressure, which is subject to considerable measurement uncertainty in practice. This work demonstrates the possibility of continuous, long-term MOC monitoring using satellite measurements.

Plain Language Summary The meridional overturning circulation (MOC) is a key ocean circulation system that moves heat, carbon, and other important tracers throughout the globe. Changes in the MOC, especially over decades or longer, greatly influence global climate. It is important to track these changes to better understand climate variability, but direct MOC measurements are logistically challenging and resource-intensive. A possible solution is using satellite data, like sea surface height, to monitor the MOC remotely. Previous research has managed to track monthly-to-yearly changes of MOC using this indirect method, but it remains unclear if this can be done over multi-year or multi-decadal periods. In this study, we demonstrate the capability of "neural networks" to achieve long-term MOC monitoring from quantities that satellites can measure, using simulations of hundreds to thousands of years of climate evolution as a test bed. Our approach also performs well near the equator, where traditional methods often fail. We additionally applied a neural network interpretation technique, which reveals that its prediction of the MOC primarily uses local and non-local information about east-to-west pressure changes, consistent with physical expectations. Our results thus provide a pathway toward accurate monitoring of the MOC using satellite data over climate-relevant timescales.

1. Introduction

The meridional overturning circulation (MOC) in the ocean is a key component of the Earth's climate system, transporting mass, heat, carbon, and other climatically crucial tracers across the oceans (e.g., Buckley & Marshall, 2016; Cessi, 2019; Lozier et al., 2019; Marshall & Speer, 2012). The critical role of the MOC in regulating the climate has been recognized for decades, with a growing body of observational and modeling

Investigation: Huaiyu Wei, Kaushik Srinivasan, Andrew L. Stewart, Aviv Solodoch, Georgy E. Manucharyan, Andrew McC. Hogg
Methodology: Huaiyu Wei, Kaushik Srinivasan, Andrew L. Stewart
Project administration: Andrew L. Stewart
Resources: Huaiyu Wei, Andrew L. Stewart
Software: Huaiyu Wei
Supervision: Kaushik Srinivasan, Andrew L. Stewart
Validation: Huaiyu Wei
Visualization: Huaiyu Wei
Writing – original draft: Huaiyu Wei
Writing – review & editing: Huaiyu Wei, Kaushik Srinivasan, Andrew L. Stewart, Aviv Solodoch, Georgy E. Manucharyan, Andrew McC. Hogg

studies emphasizing the importance of multi-decadal MOC variability in driving long-term climate changes (e.g., Zhang et al., 2019).

The significance of MOC has led to the establishment of direct MOC observations via in situ monitoring arrays in the Atlantic (Frajka-Williams et al., 2019), including the RAPID array at 26.5°N (Cunningham et al., 2007), the OSNAP array near 56°N (Lozier et al., 2017), the SAMBA array at 34.5°S (Ansorge et al., 2014), and the MOVE array at 16°N (Send et al., 2011). However, these direct observations in the Atlantic remain sparse and geographically confined. Moreover, there is no MOC monitoring array in the Southern Ocean due to logistical challenges (Meredith et al., 2013), despite its pivotal influence on global circulation, climate, carbon uptake, biogeochemical cycles, and the stability of the Antarctic ice sheet (Bennetts et al., 2024; Frölicher et al., 2015; Rintoul, 2018). Therefore, a critical gap persists in comprehensively monitoring MOC across different latitudes and ocean basins.

A potential approach to addressing this observational gap is through indirect methods that estimate MOC transport using satellite-based measurements, which provide comprehensive spatial coverage of the global ocean. The satellite-measurable quantities include ocean bottom pressure (p_b) from the GRACE mission (Tapley et al., 2004), sea surface height (η) from altimetry missions (e.g., Fu et al., 1994, 2024; Lambin et al., 2010), surface wind stress from Scatterometer missions (e.g., Figa-Saldaña et al., 2002; Hoffman & Leidner, 2005), sea surface temperature (θ_s) from the MODIS mission (Salomonson et al., 1989), and sea surface salinity (S_s) from the SMOS mission (Kerr et al., 2001).

It is theoretically justified to use these satellite-measurable quantities to deduce parts of the MOC variability. For instance, zonal differences in p_b can indicate the meridional geostrophic transport at abyssal depths, although this is complicated by the presence of bathymetric variations smaller than the resolution of satellite measurements (e.g., zonally sloping seafloors; Mazloff & Boening, 2016). Sea surface height allows for direct derivation of geostrophic currents at the surface (Little et al., 2019), and it has also been shown to negatively correlate with dense water formation, and thus deep overturning strength (Auger et al., 2024; Morrison et al., 2020; Stewart & Hogg, 2017; Yeager, 2020; Yeager et al., 2021). Zonally-averaged zonal wind stress $\left(\tau_x^{\mathrm{xavg}}\right)$ directly reflects the Ekman transport component of the MOC (McCarthy et al., 2020). Changes in the wind stress can also indirectly influence θ_s and S_s, which subsequently affect water mass transformation and MOC strength (Bishop et al., 2016). Additionally, θ_s anomalies over the subpolar North Atlantic have been used as a proxy of Atlantic MOC (AMOC) intensity, as the AMOC is responsible for the northward heat transport from the equatorial region to the North Atlantic (Caesar et al., 2018), although the efficacy of θ_s-based proxies has been debated (Little et al., 2020; Mackay et al., 2024).

Motivated by the physical relationship between the MOC and satellite-measurable quantities, considerable efforts have been dedicated to developing dynamically-based methods to reconstruct MOC variability. A prominent approach relies on the geostrophic argument that zonally integrated meridional geostrophic transport at a given latitude and depth can be determined from the p_b difference between the eastern and western boundaries of the basin. For instance, Bingham and Hughes (2008) successfully reconstructed over 90% of the inter-annual AMOC variability, defined in depth space, at 42°N in a numerical simulation. Their work also demonstrates that this reconstruction can be achieved using p_b from the western boundary alone, as variations at the eastern boundary contribute negligibly. Numerical model-based reconstructions have also been presented in subsequent studies (Bentel et al., 2015; Bingham & Hughes, 2009; Elipot et al., 2014; Hughes et al., 2018; Roussenov et al., 2008). Moreover, based on the same geostrophic argument, Landerer et al. (2015) reconstructed the observed AMOC at 26.5°N, averaged over the 3,000–5,000 m depth interval, using p_b inferred from GRACE satellite observations. Their reconstruction covered the period from 2003 to 2014, and yielded a correlation of 0.7. However, since northward and southward overturning transport at different density layers can occur at the same depth (Zhang, 2010), the MOC defined in depth space may overlook a significant portion of MOC tracer transport. Further investigations are needed to determine whether the p_b difference between the eastern and western boundaries can be used to reconstruct the MOC in density space (c.f. Section 2.2). This is particularly important at high latitudes, where the differences between these two definitions are more pronounced.

In the Southern Ocean, Mazloff and Boening (2016) demonstrated that p_b observed by the GRACE mission could account for 86% of the transport variance of Antarctic Bottom Water (AABW), which forms the lower limb of the global MOC. This relationship was identified across a section of the Pacific Ocean in the Southern Ocean State

Estimate reanalysis (Mazloff et al., 2010). More recently, Jeffree et al. (2024) systematically investigated the reconstruction of AABW transport using the ACCESS-OM2-01 coupled sea-ice/ocean model (Kiss et al., 2020). They showed that optimized linear combinations of p_b explain over 60% of the AABW transport variance at selected latitudes, even when the simulated p_b was coarse-grained to the GRACE satellite resolution of approximately 300 km. Nonetheless, they noted that significant uncertainty arises when considering realistic satellite measurement errors of p_b.

In addition to p_b, other aforementioned satellite-measurable quantities have also been employed as proxies to reconstruct MOC variability (Duchez et al., 2016; Frajka-Williams, 2015; Sanchez-Franks et al., 2021; Stewart et al., 2021; Willis, 2010). For instance, Stewart et al. (2021) demonstrated that τ_x^{xavg} is dynamically connected to the AABW export over time scales shorter than approximately 2 years in the ECCOV4r3 state estimate (Forget et al., 2015). Sanchez-Franks et al. (2021) reconstructed the geostrophic component of the AMOC in the interior Atlantic by applying the geostrophic balance to satellite-derived η, and assuming a vertical structure of the geostrophic transport based on the climatological baroclinic modes. Their AMOC reconstructions (also using Florida Strait cable observations and satellite-derived wind stress data) explained ∼70% of the observed AMOC variability at 26.5°N on inter-annual timescales.

It is important to note that while these dynamically-based methods provide valuable insights into the processes driving MOC variability, their validation has been limited to specific latitudes and ocean basins. However, AMOC variability, for example, can be latitude-dependent and gyre-specific (Lozier et al., 2010), with distinct characteristics found north and south of approximately 40°N (Bingham et al., 2007). This latitude dependence raises questions about whether the same dynamically-based methods can be effectively extended to reconstruct MOC variability in other non-validated latitudes and basins.

Recent advances in data-driven approaches have shown potential for reconstructing and monitoring the MOC across various latitudes and basins using satellite-measurable quantities. For example, Solodoch et al. (2023) demonstrated that even a simple fully-connected feed-forward neural network with a minimal number of neurons could capture 80% of the MOC strength variance across various latitudes in the Southern Ocean, Indo-Pacific Ocean, and Atlantic Ocean, though the equatorial region remained an exception. Despite these promising results, the limited temporal scope (24 years) of the ECCOV4r3 state estimate data set, used by Solodoch et al. (2023), raises concerns about the ability of this approach to capture MOC variability over decadal and longer timescales. Additionally, the ECCO state estimate cannot resolve mesoscale eddies, which feature a spatial scale of tens to hundreds of kilometers, because of the coarse model grid spacing (e.g., 100 km) adopted to reduce the computational cost. More recently, Meng et al. (2024) highlighted the advantages of using convolutional neural networks (CNN) to better account for the impacts of mesoscale eddies on MOC variability. However, their usage of idealized simulations limits the direct application of their findings to more realistic ocean basins, which feature much greater complexity in seafloor geometry, water mass structure, and atmospheric forcing (Meng et al., 2024). The superior performance of CNNs is consistent with the findings of Michel et al. (2023), who showed that a deep CNN could reconstruct the simulated historical AMOC strength at 26.5°N from North Atlantic θ_s fields with a correlation of up to 0.77, and the reconstruction is in good agreement with observations from the RAPID arrays.

In summary, several gaps remain in utilizing satellite-measurable quantities as proxies for reconstructing MOCs. First, most previous studies have focused on inter-annual MOC variability, with less attention given to reconstructing MOC variability on decadal and longer timescales. Over these longer timescales, changes in ocean stratification may alter the relationship between satellite-measurable quantities and MOC variability, making the reconstruction more challenging. Second, to the authors' knowledge, there has yet to be a successful reconstruction of the AMOC in the equatorial region, primarily due to the breakdown of geostrophic balance near the equator. Moreover, while many studies reduce the complexity of MOCs by focusing on the "MOC strength" or "maximum overturning," this simplification may overlook the temporal variations in its vertical structure (e.g., Chang & Jansen, 2022) and the complex interactions of MOC anomalies across different density levels (e.g., Yeager et al., 2021). Therefore, beyond reconstructing the MOC strength, it is also crucial to accurately reconstruct the density-dependent MOC variability.

In this study, we address these gaps by adopting a data-driven approach similar to that of Solodoch et al. (2023), but we use two pre-industrial (PI) control simulations from the Coupled Model Intercomparison Project Phase 6 (CMIP6). These simulations, which span hundreds to thousands of model years, allow us to investigate the

reconstruction of intrinsic MOC variability on decadal and longer timescales. We propose a non-local dual-branch neural network (DBNN) architecture, which demonstrates promising capabilities in capturing the variability of the AMOC, the Southern Ocean mid-depth MOC, and the Southern Ocean abyssal MOC across sub-annual to multi-decadal timescales. Notably, the proposed neural network framework successfully reconstructs both the time-varying strength and vertical structure of the MOC, providing a more complete solution for monitoring MOC variability.

The remainder of this article is organized as follows: Section 2 introduces the two CMIP6 simulations used in this study and presents the proposed neural network architecture. In Section 3, we demonstrate the skill of the proposed neural network in reconstructing MOC strengths and examine the utility of various satellite-measurable quantities, using one of the CMIP6 simulations. Section 4 investigates the underlying physics learned by the neural network using Layer-wise Relevance Propagation, a technique that interprets neural networks by tracing predictions back through the layers to the input variables. In Section 5, we extend the neural network's application from MOC strength reconstruction to reconstructing MOC variability at various density levels. Section 6 evaluates the robustness of the proposed neural network using data from a different CMIP6 simulation with an eddy-permitting resolution. Concluding remarks and discussion follow in Section 7.

2. Methods

To develop a practical approach for reconstructing MOC variability, the ideal method would use direct MOC observations and satellite-measurable quantities from the real ocean. However, the scarcity of in situ MOC observations, especially over longer timescales, limits this pathway. At the current stage, our goal is to demonstrate that the long-term MOC variability is inherently reflected in satellite-measurable quantities and can be captured using a data-driven approach, instead of aiming to propose a practical MOC monitoring solution. We therefore develop and test our data-driven method using outputs from CMIP6 climate simulations. In the following subsections, we describe the climate simulations used, our definition of MOC variability, the proposed neural network architecture, and the training process for the neural network.

2.1. PI Control Simulations

We start by using the PI control simulation from ACCESS-ESM1.5 (Ziehn et al., 2020) to develop and validate our machine learning approach. The ocean component of ACCESS-ESM1.5 employs Modular Ocean Model (MOM) version 5 (Griffies, 2012), coupled with the CICE4.1 sea-ice model (Hunke et al., 2010). The ocean model has a nominal horizontal resolution of 1°, with finer resolution near the equator (0.33°) and over the Southern Ocean (~0.4° at 70°S). A flow-dependent eddy buoyancy diffusivity (Ferrari et al., 2010; Griffies et al., 2005) is adopted to parameterize the eddy-induced overturning circulation (Gent & McWilliams, 1990; Gent et al., 1995). The PI control run spans 1,000 years, and monthly outputs are used in this study. Both the satellite-measurable oceanic quantities (i.e., p_b, η, θ_s, and S_s) and MOC streamfunction (c.f. Section 2.2) on the model's native ocean grid are used, comprising 300 latitudes and 360 longitudes. Zonal wind stress τ_x is linearly interpolated from the native atmospheric grid to the native ocean grid.

To further assess the robustness of our machine learning approach across models, we also use the PI control run from GFDL-CM4 (Held et al., 2019). The ocean component of GFDL-CM4 is based on MOM6, with a nominal horizontal grid spacing of 0.25° (Adcroft et al., 2019). This finer resolution allows for partially resolved mesoscale eddies, and better resolves basin boundaries and bathymetry, both of which help to improve the simulated western boundary current and MOCs. Unlike ACCESS-ESM1.5, no advective or diffusive mesoscale eddy closure is adopted. The GFDL-CM4 PI control run spans 500 years, and monthly outputs are used in this study. While the MOC streamfunction on the native latitude grid is adopted, the oceanic satellite-measurable quantities interpolated to a 1° resolution grid are used in constructing the neural network, as the publicly available outputs of p_b and η are only available on this grid. Zonal wind stress is linearly interpolated from the atmospheric grid to the interpolated ocean grid.

2.2. MOC Strength Definition

In this study, we account for three components of the MOC: the Southern Ocean (SO) abyssal MOC, the SO mid-depth MOC, and the mid-depth Atlantic MOC (AMOC; excluding the Mediterranean Sea). The latitude ranges of

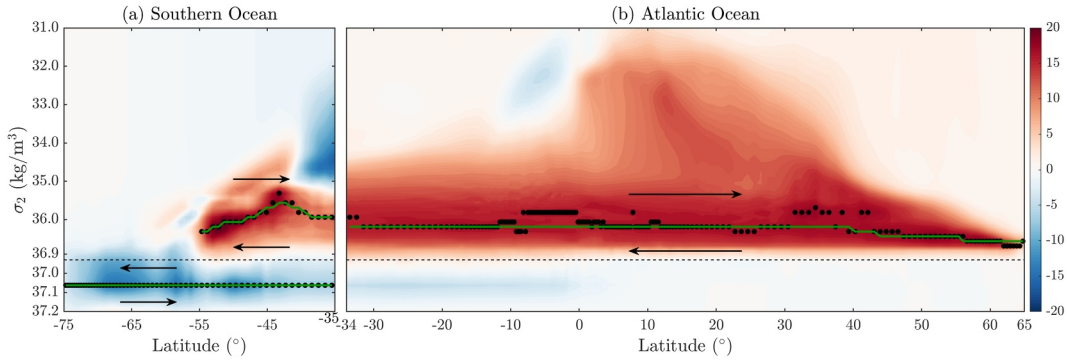

Figure 1. Density-space meridional overturning circulation (MOC) averaged in the ACCESS-ESM1.5 pre-industrial (PI) control run over the (a) Southern Ocean (SO) and (b) the Atlantic Ocean, expressed in units of Sverdrups (1 Sv $\equiv 10^6$ m^3/s). The circulation follows streamlines in the directions indicated by the black arrows. Black dots indicate the potential density level where the time-mean overturning magnitude is maximized at each latitude for the SO mid-depth MOC, the SO abyssal MOC, or the mid-depth Atlantic MOC (AMOC). The potential density 36.93 kg/m^3 separating the mid-depth and abyssal MOCs is indicated by the black dashed line. The green line represents the density level at which the MOC strength is defined in Section 2.2.

these MOC cells considered in this study are [75°S, 35°S], [55°S, 35°S], and [34°S, 65°N], respectively. Each MOC is quantified using the streamfunction calculated in density space, defined as

$$\Psi(y,\,\sigma_2,\,t) = -\int_{x_w}^{x_e} \mathrm{d}x \int_{\eta_b(x,y)}^{\eta_{\sigma_2}(x,y,t)} v_r \,\mathrm{d}z, \tag{1}$$

where σ_2 is the potential density anomaly referenced to 2 km depth (minus 1,000 kg m^{-3}), x is longitude, y is latitude, and v_r indicates the residual meridional velocity that contains both the Eulerian velocity and the parameterized eddy-induced velocities, if applicable (Griffies et al., 2016). The zonal integral is performed across the corresponding ocean basin from the western endpoint x_w to the eastern endpoint x_e ($\int_{x_w}^{x_e}$ becomes $\oint$ for SO MOCs), and the vertical integral is performed from the seafloor η_b to the isopycnal depth η_{σ_2}. We derive $\Psi(y, \sigma_2, t)$ directly from the monthly CMIP6 variables "msftmrho" or "msftyrho," normalized by the Boussinesq reference density. The time averages of "msftmrho" and "msftyrho" are accumulated at each model time step according to the diagnostic protocol (see Sections 6 of Griffies et al., 2016). Figure 1 shows the time mean streamfunction for each of these components of the MOC, averaged over the entire ACCESS-ESM1.5 PI control run, with clockwise (anti-clockwise) flow around positive (negative) values. The time mean structure of the MOC in the PI control simulation of GFDL-CM4 is shown in Supporting Information S1 (Figure S1).

The majority of our analyses focus on reconstructing the strength of these three MOCs at each latitude following Solodoch et al. (2023). The reconstruction of density-dependent MOC (i.e., both its latitudinal and vertical structures) is deferred to Section 5. To define the strength of the MOCs, we first identify the potential density level where the time-mean overturning magnitude reaches its maximum at a given latitude for each MOC (indicated by black dots in Figure 1), which can be specified as

$$\sigma_2^*(y) = \arg\max_{\sigma_2}\left\{\left|\overline{\Psi}^t(y,\sigma_2)\right|\right\}. \tag{2}$$

Here, $\overline{\bullet}^t$ denotes the time-mean operator across the entire PI control simulation, and the arg max is taken over a subset of the full density range that corresponds to the specific overturning cell of interest (either the mid-depth or abyssal cell; the separation density level $\sigma_2 = 36.93$ kg/m^3 is shown by the black dashed line in Figure 1). Notably, σ_2^* exhibits non-monotonic variations with latitude in the Atlantic (e.g., around 10°S and 40°N). We thus opt to define the MOC strength as

$$\psi(y,t) = \Psi\left(y,\,\sigma_2^\dagger(y),\,t\right), \tag{3}$$

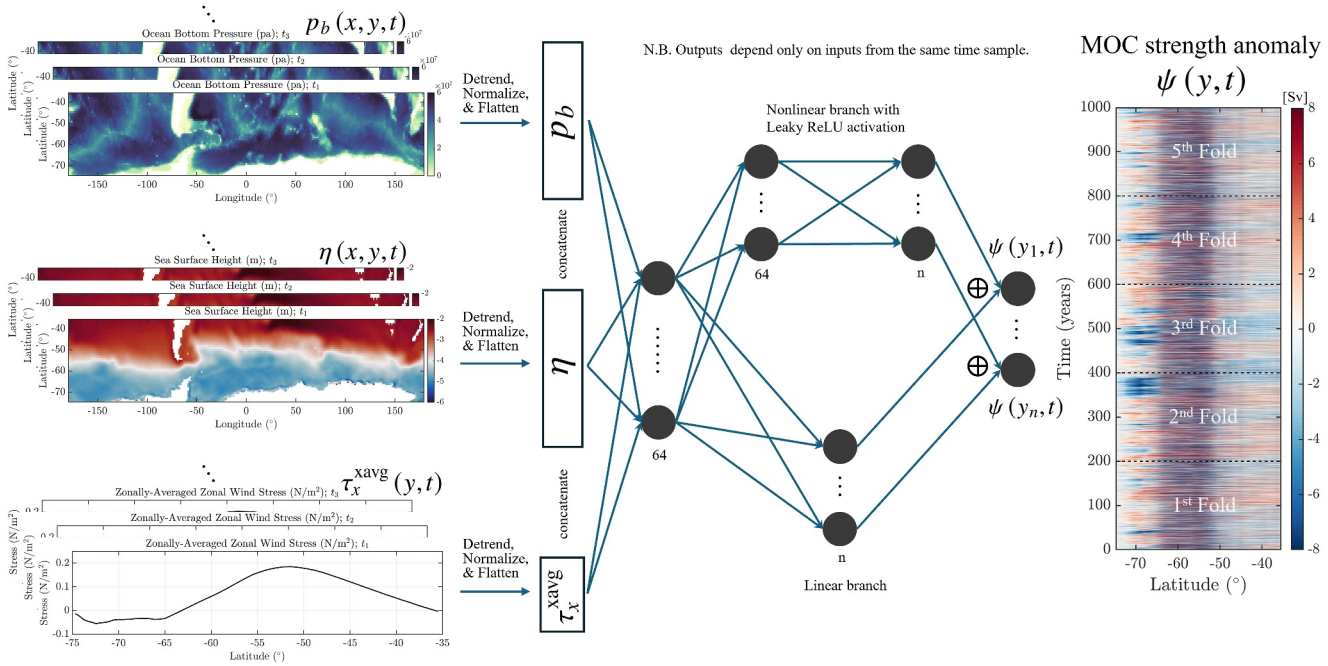

Figure 2. Schematic showing the architecture of the proposed baseline non-local dual-branch neural network (DBNN). On the left, subplots illustrate examples of satellite-measurable quantities from the entire MOC basin, including ocean bottom pressure p_b (top), sea surface height η (middle), and zonally-averaged zonal wind stress τ_x^{xavg} (bottom). These quantities are detrended, normalized, and flattened before being concatenated into the input layer of the DBNN. The network includes a dual-branch hidden layer: a nonlinear branch using Leaky ReLU activation (top) and a linear branch without activation (bottom). The output layer, formed by the summation of these branches, is designed to reconstruct the temporal variations in the detrended MOC strength anomalies across various latitudes, as shown in the rightmost subplot. While the subplots in this schematic use data from the SO abyssal MOC, the same architecture applies to the SO mid-depth MOC and the AMOC.

where Ψ is the time-varying overturning streamfunction, and the density level $\sigma_2^\dagger(y)$ is obtained by selecting the most common value of $\sigma_2^*(y)$ south of 5°N for the AMOC, followed by smoothing σ_2^* across all three MOCs using local regression (Cleveland, 1979). This specific definition of $\sigma_2^\dagger$ removes the non-monotonic variations in σ_2^* south of 5°N in the AMOC for the ACCESS-ESM1.5 simulation, partly motivated by the absence of such variations in the GFDL-CM4 simulation (Figure S1 in Supporting Information S1). The resulting $\sigma_2^\dagger(y)$, shown by the green line in Figure 1, follows a path through the core of the MOC. Defining the MOC strength via σ_2^* instead of $\sigma_2^\dagger$ yields a slightly less accurate, but qualitatively similar, reconstruction using the proposed machine learning method. To ensure that a positive ψ anomaly always corresponds to an enhanced MOC, a negative sign is added in defining the strength of the SO abyssal MOC (whose time-mean streamfunction is negative), that is, $\psi(y,t) = -\Psi(y, \sigma_2^\dagger, t)$.

In Section 5, we discuss the reconstruction of density-dependent MOC, $\Psi(y, \sigma_2, t)$, which circumvents the limitations of defining a single MOC strength measure at each latitude.

2.3. Machine Learning Methodology

2.3.1. Neural Network Architecture

We propose a non-local dual-branch neural network (DBNN) to learn the relationship between MOC strength anomalies at various latitudes and satellite-measurable quantities across the entire corresponding MOC basin. We use input variables and MOC strength that are in-phase to construct the neural network (i.e., without any temporal memory), framing our methodology as a regression problem rather than a time series forecasting problem (e.g., Zhai et al., 2024) within the machine learning context. Figure 2 illustrates our baseline DBNN architecture, which uses ocean bottom pressure p_b, sea surface height η, and zonally-averaged zonal wind stress τ_x^{xavg} as input variables to reconstruct the variability of the SO abyssal MOC. This specific input combination was chosen due to its skill in MOC reconstruction across different basins and timescales. Other input combinations of satellite-measurable quantities are discussed in Section 3.2. We use zonally-averaged instead of zonally-varying zonal

wind stress, following Solodoch et al. (2023), because it is more directly related to the total meridional Ekman flux (e.g., Vallis, 2017), and to high-frequency abyssal export in the Southern Ocean (Stewart et al., 2021). The satellite-measurable quantities from the entire MOC basin are flattened and concatenated into the input layer. Throughout this article, we will refer to the satellite-measurable quantities used by the DBNN as input variables, and the elements of the input array (i.e., the input variables at specific geographic locations) as input features.

The input layer is fully connected to the first hidden layer, which consists of 64 neurons, followed by two branches. The first branch, referred to as the non-linear branch, applies the Leaky Rectified Linear Unit (ReLU) activation function (e.g., Manucharyan et al., 2021; Xie et al., 2023a),

$$\sigma(x) = \begin{cases} x, & \text{if } x > 0, \\ 0.2x, & \text{if } x \leq 0. \end{cases} \tag{4}$$

The second branch, referred to as the linear branch, has no activation function applied. The non-linear branch consists of two hidden layers with 64 and n neurons, where n represents the number of latitudes in the MOC of interest, while the linear branch includes only one hidden layer with n neurons. The output layer, formed by summing the outputs from both branches, is designed to reconstruct the temporal variations of the MOC strength anomalies across latitudes (rightmost subplot in Figure 2). The mathematical formulation of the proposed DBNN is given in Appendix A.

Both the input features and MOC strength at each latitude are detrended and normalized. The normalization is achieved by removing the mean and dividing by the standard deviation calculated in the training data set. The normalization addresses the magnitude differences among input variables, while the detrending is to remove the linear drift in η and thus p_b found in the ACCESS-ESM1.5 PI control run (Figure S2 in Supporting Information S1). The dual-branch structure is inspired by the ResNet architecture (He et al., 2016), and is motivated by the fact that linear geostrophic balance and Ekman dynamics play a zero-order role in constraining large-scale ocean flows, whereas ocean dynamics is inherently non-linear and chaotic (e.g., Vallis, 2017). The dual-branch structure thus enables the neural network to simultaneously infer linear and non-linear relationships for reconstructing MOCs. In Supporting Information S1 (Figure S5), we demonstrate that a linear neural network achieves comparable skill in reconstructing MOC strength, though its performance varies more strongly across different latitudes.

Notice that the DBNN is capable of learning latitudinally non-local relationships between MOC variability and satellite-measurable quantities, as both the inputs and outputs of the DBNN span multiple latitudes across the domain. The rationale for incorporating spatially non-local information is twofold: First, non-local connections between the input variables and MOC variability are theoretically motivated, for example, high-latitude surface buoyancy has been shown to be connected to low-latitude MOC variability (Cessi, 2019; Sun et al., 2020). Second, MOC variability has been shown to exhibit coherence over certain latitude ranges, especially on decadal and longer timescales (Bingham et al., 2007; Elipot et al., 2017; Gu et al., 2020; Wunsch & Heimbach, 2013). This suggests that reconstructing the MOC at one latitude may aid in reconstructing MOC at nearby latitudes. In Section 3.3, we will demonstrate that the capability of accounting for non-local information in the neural network, combined with the coherence of the AMOC, is critical for reconstructing AMOC variability near the equatorial region, particularly for the long-term variability.

2.3.2. Neural Network Training

The DBNN is trained using the Adam optimization algorithm (Kingma & Ba, 2014). The training process minimizes the mean squared error loss function, which quantifies the squared mismatch between the reconstructed and diagnosed MOC strength. To mitigate overfitting, we apply L2 regularization (George et al., 2021; Hoerl & Kennard, 1970) to the loss function, using a weight decay coefficient of 0.01. The L2 regularization introduces a penalty term that limits the squared norm of the weight parameters (i.e., favoring smaller weights) (Krogh & Hertz, 1991; McDonald, 2009). Additionally, a dropout layer with a dropout rate of 0.2 is applied after each hidden layer to further prevent overfitting during training (Srivastava et al., 2014). The learning rate and batch size are set to 0.001 and 600, respectively. Early stopping is employed by monitoring the testing loss (Srinivasan et al., 2024), and the training process is terminated if no improvement has been observed for 50

epochs (Xie et al., 2023b). We select the trained neural network with the weakest testing loss over the entire duration of training.

To improve the robustness of the training, we employ 5-fold cross-validation (Michel et al., 2023; Stone, 1974), which divides the data set into five subsets, or folds. Each fold serves as the testing data set once, while the remaining folds are used to train the neural network. For example, the 1000-year monthly data from the ACCESS-ESM1.5 PI control run is partitioned into five folds, each containing 200 years of data (as shown in the rightmost panel of Figure 2). During the cross-validation, DBNN is trained five times, with each fold used as the testing data set once. The reconstructed MOC is obtained by combining the predictions from the testing data sets of all five trained DBNNs each covering 200 years. To further enhance robustness, we employ ensemble training, since the neural network's performance depends on the initial weights (e.g., Solodoch et al., 2023). The 5-fold cross-validation training process is repeated five times with different random initial weights, resulting in a total of 25 neural networks trained for a single MOC reconstruction. The final MOC strength reconstruction is obtained by averaging the predictions from these five ensembles. The epistemic uncertainty is evaluated by examining the spread of reconstruction skill across the five ensembles. This spread is generally negligible (c.f. Figure 4), likely due to the multiple regularization strategies applied during training. Note that the reconstruction skill of the ensemble-averaged prediction is not necessarily confined within the range of reconstruction skill achieved by individual ensemble members.

2.3.3. Low-Pass Filter

A key focus of this study is the reconstruction of long-term MOC variability. To facilitate this, we apply a temporal low-pass filter (LPF) to the time series of both the input variables at each geographic location and MOC strength at each latitude prior to neural network training. The LPF is implemented as a fifth-order Butterworth filter, and it only uses data from the past to ensure that the filter is causal. To minimize filter boundary effects, the time series is padded at the beginning using reflective padding, which mirrors the data. Two cutoff timescales, 2 and 10 years, are used in this study to emphasize the inter-annual and multi-decadal variability of the MOC, respectively. The Hovmöller diagrams showing the full MOC variability and the long-term MOC variability (i.e., with the 10-year LPF) in ACCESS-ESM1.5 are presented in Figures 3a and 3d, respectively.

As detailed later in Section 3.1 (c.f. Figure 4), we note that the optimal approach for reconstructing decadal and longer MOC variability is to train the DBNN using input features and MOC strengths that were filtered with the 2-year LPF, and subsequently applying the 10-year LPF to the predicted MOC strengths. Thus, all long-term MOC reconstructions presented in this manuscript follow this approach unless explicitly stated otherwise.

2.3.4. Skill Metric

We quantify the reconstruction skill of the DBNN by calculating the Nash and Sutcliffe (1970) coefficient of efficiency, which is also commonly referred to as the coefficient of determination in the machine learning community, that is,

$$R^2 = 1 - \frac{\sum_{i=1}^{N_{\text{samp}}} \left(\psi - \psi_{\text{rec}}\right)^2}{\sum_{i=1}^{N_{\text{samp}}} \left(\psi - \overline{\psi}^{N_{\text{samp}}}\right)^2}. \tag{5}$$

Here N_{samp} stands for the total number of the monthly data, $\overline{\bullet}^{N_{\text{samp}}}$ indicates the arithmetic mean operator across all samples, ψ and ψ_{rec} represent the diagnosed and reconstructed MOC strength, respectively. One should note that R^2 thus defined is not equivalent to the squared Pearson correlation coefficient. The R^2 value ranges from $-\infty$ to 1, where $R^2 = 1$ implies a perfect fit (i.e., $\psi = \psi_{\text{rec}}$), and a negative R^2 indicates that the reconstructed MOC strength performs worse than the arithmetic mean of the diagnosed ψ (i.e., $\overline{\psi}^{N_{\text{samp}}}$). Additionally, R^2 is equivalent to the explained variance metric when there is no systematic bias in the prediction; otherwise, R^2 is lower.

3. MOC Strength Reconstruction in the ACCESS-ESM1.5 PI Control Run

In this section, we first demonstrate the skill of our baseline DBNN in reconstructing MOC variability across different timescales. Next, we assess the importance of each input variable of the baseline DBNN (ocean bottom pressure p_b, sea surface height η, and zonally-averaged zonal wind stress τ_x^{xavg}) by analyzing the performance of

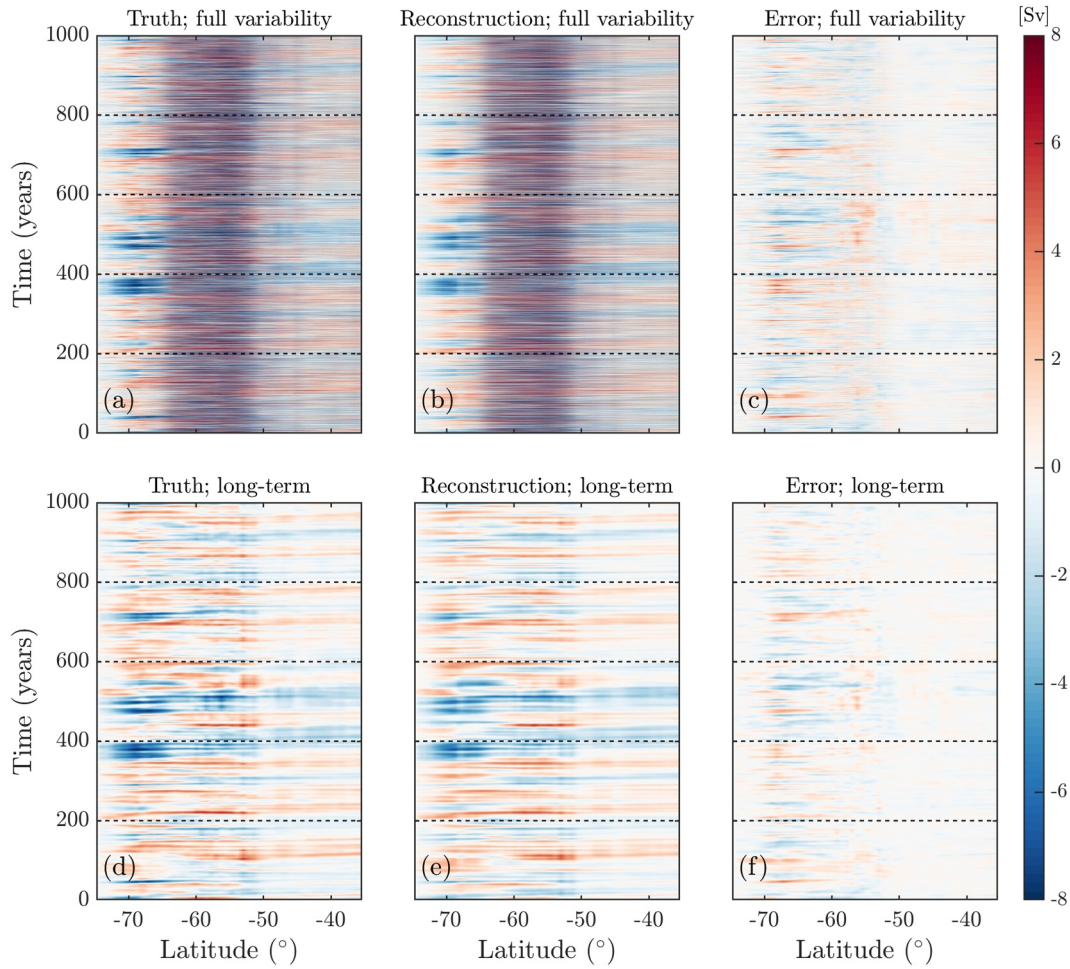

Figure 3. Hovmöller diagram of the SO abyssal MOC strength anomalies diagnosed from the ACCESS-ESM1.5 PI control run (left column), reconstructed via the DBNN described in Section 2.3 and Figure 2 (middle column), and the reconstruction error (right column; reconstructed minus diagnosed). Panels (a)–(c) show the full variability of MOC strength anomalies, while panels (d)–(f) present the variability on decadal and longer timescales.

the DBNN using subsets of these variables. Additionally, we investigate whether incorporating additional information of sea surface temperature (θ_s) and sea surface salinity (S_s) enhances the reconstruction skill of the DBNN.

3.1. MOC Strength Reconstruction Skill of the Baseline DBNN

In Figure 3, we compare the diagnosed SO abyssal MOC strength anomalies from the ACCESS-ESM1.5 PI control run against the reconstructed MOC strength anomalies by the baseline DBNN. Unless specified otherwise, the MOC anomaly indicates the deviation from the temporal mean MOC strength and has been detrended. Without applying the LPF (upper panels in Figure 3), strong seasonal variability dominates the total variability, particularly between ∼65°S and ∼50°S, where the westerly winds have strong seasonality. The pronounced seasonality of the SO abyssal MOC aligns with the finding from Stewart et al. (2021), which shows that Southern Ocean winds drive short-term variability in the SO abyssal MOC. The 10-year LPF extracts decadal and longer variability (lower panels in Figure 3), with the largest anomalies exceeding 6 Sverdrup at 68°S during years 350–400. Northward propagation of these longer-term variabilities is evident in Figure 3b, consistent with the primary role of dense water formation at high latitudes in driving SO abyssal MOC variability and the northward export of dense water (e.g., Kusahara et al., 2017; Solodoch et al., 2022; Stewart & Hogg, 2017).

Table 1
Summary of Time-Mean Absolute Error (MAE), Time Standard Deviation (MSD), and Normalized Error MAE/MSD, Averaged Across Latitudes for Different MOCs and Variability Timescales in the ACCESS-ESM1.5 PI Control Run

MOC	Variability timescales	MAE (Sv)	MSD (Sv)	MAE/MSD (%)
SO abyssal MOC	Full variability	1.03	5.08	23.2
	Inter-annual and longer	0.57	2.09	27.6
	Decadal and longer	0.41	1.49	26.4
SO middepth MOC	Full variability	1.18	8.30	14.8
	Inter-annual and longer	0.55	2.45	22.6
	Decadal and longer	0.32	1.09	29.4
AMOC	Full variability	0.81	2.85	28.1
	Inter-annual and longer	0.38	1.20	31.0
	Decadal and longer	0.20	0.78	26.4

The reconstruction error is generally much smaller in magnitude than the diagnosed MOC strength anomalies across most time samples and latitudes. Notably, the DBNN successfully captures the northward propagation signal, even though the DBNN only uses in-phase information (i.e., without any information back in time), indicating that these propagating MOC anomalies can be well represented by variations in the input variables (e.g., dense water anomalies alter p_b). In addition, we note that extreme events are underestimated by the DBNN, such as the significant negative anomaly around 68°S during years 350–400.

In Supporting Information S1 (Figures S3 and S4), we present Hovmöller diagrams of the diagnosed and reconstructed MOC strength anomalies, as well as the corresponding reconstruction errors, for the SO mid-depth MOC and the AMOC. They all demonstrate the promising reconstruction skill of the proposed baseline DBNN. We also summarize in Table 1 the time-mean absolute reconstruction error averaged across all latitudes for each MOC, showing that the absolute error typically remains below 30% of the standard deviation of the diagnosed MOC variability, across all MOCs and timescales considered.

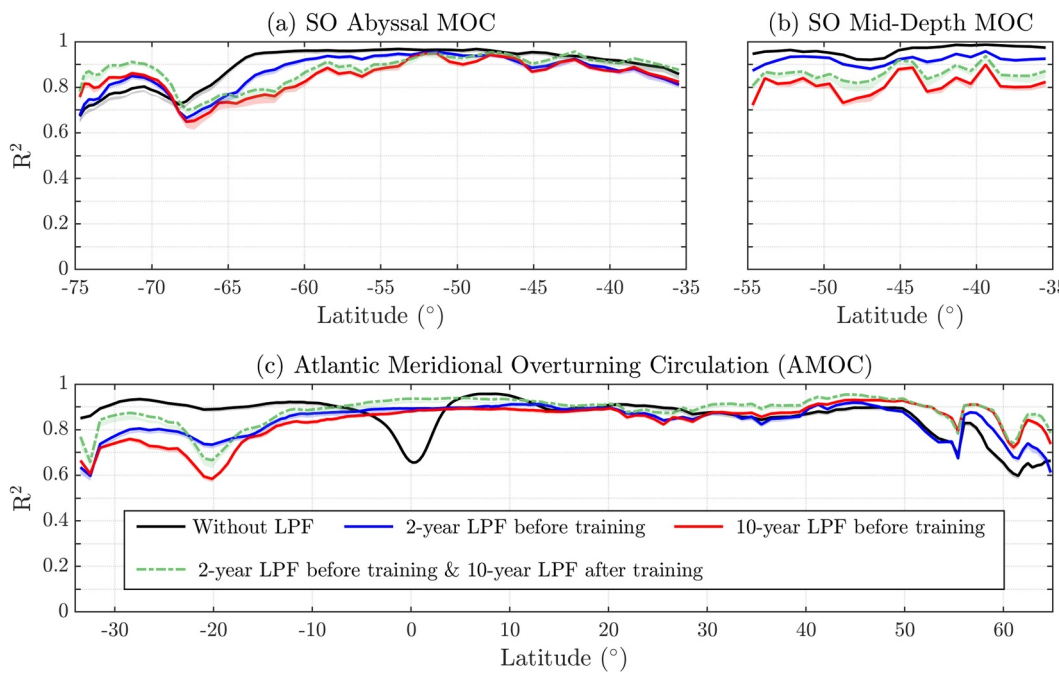

Figure 4. Reconstruction skill [R^2 defined via Equation 5] of the baseline DBNN in the ACCESS-ESM1.5 PI control run for the (a) SO abyssal MOC, (b) SO mid-depth MOC, and (c) AMOC. The shadings illustrate the spread of reconstruction skill across the five ensembles.

To quantitatively evaluate the MOC reconstruction by the baseline DBNN at each latitude, we present in Figure 4 the coefficient of efficiency R^2 defined via Equation 5 for the three MOCs of interest. In general, R^2 exceeds 0.6 across almost all latitudes of the three MOCs regardless of whether LPF is applied or not, and reaches values as high as 0.9 in certain regions (e.g., the northern part of the SO abyssal MOC, the SO mid-depth MOC without LPF or with the 2-year LPF, and the subtropical AMOC in the Northern Hemisphere). Importantly, AMOC variability in the equatorial region, where the local geostrophic balance does not hold, is well reconstructed on inter-annual and longer timescales ($R^2 \sim 0.9$), though the skill drops to $R^2 \sim 0.7$ when including the sub-annual variability (i.e., black line). We will discuss in Section 3.3 the reasons for the success of DBNN in reconstructing equatorial AMOC variability and the skill drop when considering the high-frequency variability.

The DBNN shows slightly reduced skill in reconstructing long-term variability of the SO MOCs compared with the reconstruction of high-frequency variability. For instance, in the SO mid-depth MOC, the mean and standard deviation of the R^2 values are 0.96 ± 0.02, 0.92 ± 0.02, and 0.81 ± 0.04 for cases without LPF, with the 2-year LPF and with the 10-year LPF applied before training, respectively. The drop in reconstruction skill for longer timescales is likely due to the higher temporal autocorrelation in the data filtered by longer-term LPF, which equivalently reduces the amount of independent data available for training the neural network.

To further assess the fidelity of the reconstructed MOC on decadal and longer timescales, we also applied the LPF to DBNN predictions after the network had been trained. The green line in Figure 4 illustrates the reconstruction skill obtained by training the DBNN using input features and MOC strengths that were filtered with a 2-year LPF, and subsequently applying a 10-year LPF to both the predicted and the corresponding diagnosed MOC strengths (i.e., 2-year and 10-year LPFs are applied before and after training, respectively). This approach yields a notably higher R^2 compared to training the DBNN directly using data with the 10-year LPF applied. Similarly, Solodoch et al. (2023) and Meng et al. (2024) reported improved MOC reconstruction by neural networks when the LPF was applied after training, whereas applying the LPF before training resulted in degraded performance. The improved DBNN performance when the LPF is applied after training is consistent with our argument that the reduced skill for decadal and longer variability when the LPF is applied before training is due to the reduced number of effective data with a longer timescale LPF applied. We also evaluated the reconstruction skill by applying the 10-year LPF to the predictions from the DBNN trained on unfiltered data, which resulted in worse performance compared to applying the LPF before training (Figure S6 in Supporting Information S1). This is likely because, without LPF, high-frequency variability dominates the full variability of the MOC strength, causing the DBNN to focus on high-frequency patterns during training, thereby reducing its ability to capture long-term variability. This issue cannot be fully resolved by deseasoning the data (i.e., subtracting the monthly climatological mean), as deseasoning does not remove all high-frequency variability that may still distract the neural network from learning the underlying long-term changes (Figure S6 in Supporting Information S1). Unless stated otherwise, the long-term MOC reconstruction presented in this study is achieved by applying the 10-year LPF to the reconstruction from the DBNN trained on data filtered with the 2-year LPF.

The AMOC reconstruction skill on inter-annual and longer timescales drops in the Southern Hemisphere. This skill reduction may be attributed to the fact that the DBNN has access to more input features from the Northern Hemisphere (0° to 65°N) than the Southern Hemisphere (34°S to 0°), due to the definition of the AMOC domain in this study (i.e., 34°S to 65°N). This imbalance suggests that there is less effective non-local information available for reconstructing the AMOC in the Southern Hemisphere compared to the Northern Hemisphere, and likely leads to the reduced skill of the DBNN in reconstructing the AMOC in the Southern Hemisphere. Reconstruction skill also declines at high latitudes in the SO abyssal MOC between 75°S and 65°S. One possible explanation for this decline is the lack of sea ice information in the input variables used by the baseline DBNN. Dense water formation predominantly occurs in open ocean polynyas in ACCESS-ESM1.5, an effect that might not be adequately represented without including sea ice information in the DBNN. However, attempts to include sea ice quantity (e.g., sea ice concentration, sea ice thickness, or sea ice mass) as an additional input variable did not improve the reconstruction skill compared to the baseline DBNN (not shown). The exact reason for the skill decline at high latitudes remains elusive to the authors, and further investigations are warranted.

We also examine how the reconstruction skill depends on the amount of training data available to the neural network, by conducting additional experiments using subsets of the full ACCESS-ESM1.5 data set to train the DBNN (Figure S7 in Supporting Information S1). The results indicate that approximately 30 years of data are

sufficient to capture the full AMOC variability, whereas substantially more data (e.g., 100–200 years) are needed to skillfully reconstruct the long-term AMOC variability.

3.2. Skill of the DBNN With Different Input Variables

In the previous section, we demonstrated the MOC strength reconstruction skill of the DBNN using p_b, η, and τ_x^{xavg} as input variables. In this section, we proceed to investigate whether adding additional input variables, namely θ_s and S_s, would improve the reconstruction skill of DBNN. This is motivated by previous studies of AMOC fingerprints (e.g., Zhang, 2008) and the dynamical link between MOC strength and high-latitude surface density (e.g., Galbraith & de Lavergne, 2019; Sun et al., 2020). Moreover, we examine various combinations of these input variables, from which we gain insights into their relative importance in constraining the MOC variability.

Since the reconstruction skill varies with latitude (e.g., Figure 4), we adopt box-whisker plots to present the performance of the DBNN with different combinations of input variables. Figure 5 illustrates the DBNN's skill in reconstructing decadal and longer timescale MOC strength variability, while the skill for reconstructing the full and inter-annual variability is provided in Supporting Information S1 (Figures S8 and S9).

In the reconstruction of long-term AMOC variability (Figure 5c), significant skill drops occur when p_b is excluded as an input variable for the DBNN (e.g., the low-limit $R^2 < 0.4$ for all tested cases without p_b). This irreplaceable importance of p_b underscores the need to improve the accuracy of p_b observations. The much lower reconstruction skill using η compared to p_b aligns with the findings of Roussenov et al. (2008) and Bingham and Hughes (2009), which suggested that p_b and η variations on inter-annual and longer timescales can be quite distinct in the Atlantic, especially in deep regions away from the shelf. While adding η and τ_x^{xavg} alongside p_b yields nearly the same reconstruction skill as using p_b alone, the skill drops slightly when all five satellite-measured quantities—p_b, η, τ_x^{xavg}, θ_s, and S_s—are included (e.g., the median R^2 drops from 0.91 to 0.88 after including θ_s and S_s). Similar skill reductions are also observed in the reconstruction of SO MOCs (Figures 5a and 5b). In theory, adding more input variables should not degrade the performance of a neural network, as a perfectly-trained network can assign near-zero weights and biases to irrelevant inputs. However, in practice, the inclusion of θ_s and S_s increases the number of trainable parameters in the DBNN from 1,360,728 to 2,683,864 for AMOC reconstruction, requiring significantly more data for proper training. Additionally, redundant features can dilute key signals (e.g., geostrophic balance), making it more challenging for the network to identify effective patterns. Therefore, we argue that in-phase θ_s and S_s are only loosely related to MOC strength, and their inclusion does not provide useful improvement beyond the information that has already been captured by p_b, η, and τ_x^{xavg}. Additionally, we explored augmenting the baseline DBNN by incorporating θ_s and S_s at various lagged times ranging from 6 months to 10 years as additional input variables, but this did not lead to improved reconstruction skill.

Although p_b remains the most effective sole predictor for reconstructing the SO mid-depth and abyssal MOCs (Figures 5a and 5b), there exist alternative input combinations with similar performance. In particular, the combination of η and τ_x^{xavg} (indicated by the gray arrow) achieves nearly the same reconstruction skill (median $R^2 > 0.8$) as input combinations that include p_b. We conjecture that DBNN with η and τ_x^{xavg} as inputs may offer a promising practical alternative for reconstructing SO MOCs, given the relatively large observational uncertainties associated with p_b (e.g., Jeffree et al., 2024). The importance of τ_x^{xavg} in reconstructing the SO mid-depth MOC indicates that wind-driven Ekman overturning is only partially compensated by the parameterized eddy-induced overturning in ACCESS-ESM1.5, which is also consistent with high-resolution eddy-resolving simulations (Abernathey et al., 2011; Bishop et al., 2016; Kong & Jansen, 2021; Mak et al., 2018; Morrison & Hogg, 2013). The moderate reconstruction skill (median $R^2 > 0.5$) achieved by S_s or θ_s suggests that variations in SO mid-depth MOC strength may leave a detectable fingerprint in the sea surface salinity, temperature, or density field, though the direct dynamical links among them are less clear.

In reconstructing the SO abyssal MOC (Figure 5a), τ_x^{xavg} alone yields very low skill (median $R^2 < 0.1$), while all other tested input combinations demonstrate good skill (median $R^2 > 0.7$). As indicated by the gray diamond (see also Figure S8 in Supporting Information S1), τ_x^{xavg} alone achieves a median R^2 of 0.61 when reconstructing the full-variability (i.e., without applying LPF) of the SO abyssal MOC strength. The inability of τ_x^{xavg} to reconstruct long-term SO abyssal MOC variability, while successfully capturing shorter-term variability, is consistent with the findings of Stewart et al. (2021), which links Southern Ocean winds to SO abyssal MOC variability on

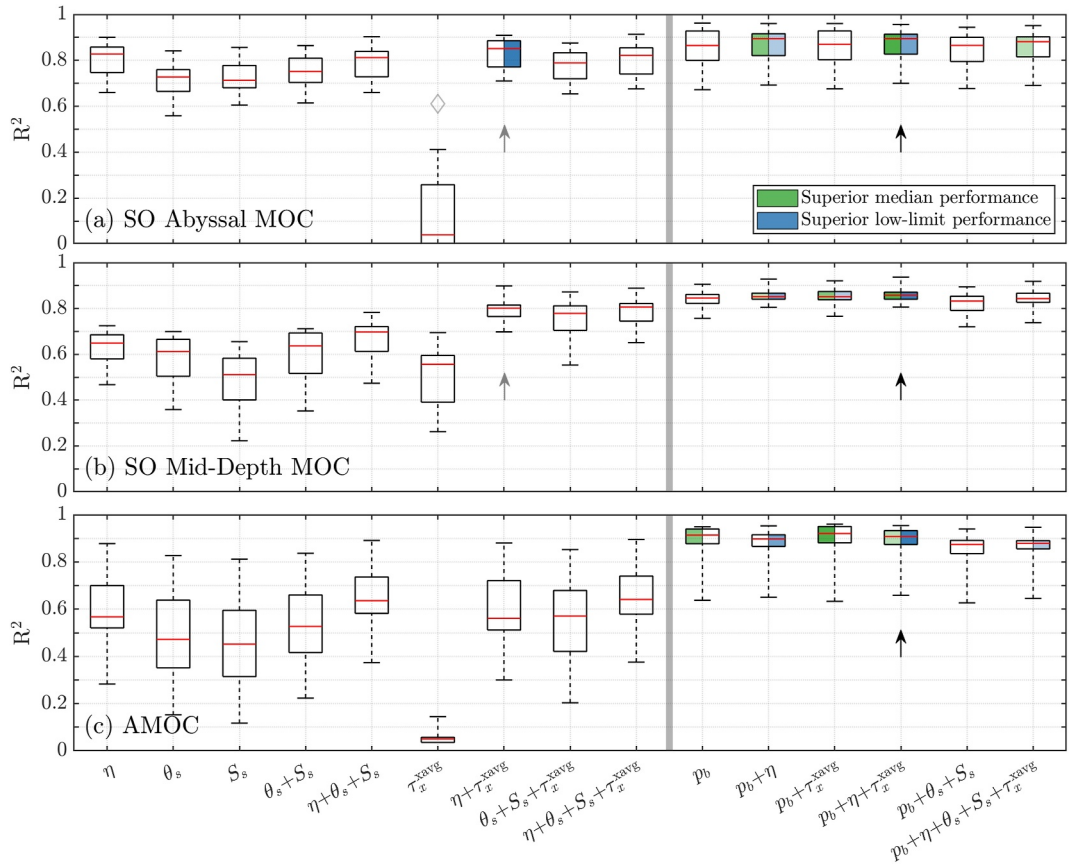

Figure 5. Box-whisker plot of the reconstruction skill [R^2 defined via Equation 5] with various combinations of input variables for the (a) SO abyssal MOC, (b) SO mid-depth MOC, and (c) AMOC variability on decadal and longer timescales. The bottom and top edges of the box indicate the first (25%) and third (75%) quartiles of the latitudinally varying R^2 values, respectively. The red line denotes the median R^2 across all latitudes. The whiskers extend to the smallest and largest R^2 values. Green and blue shadings highlight input variable combinations that have superior median and low-limit performance, respectively. The gray line separates input variable combinations with or without p_b. The input variable combination of our baseline DBNN is highlighted by the black arrow. An alternative input variable combination that works well in reconstructing SO MOCs are indicated by the gray arrow. The gray diamond in panel (a) shows the median R^2 for reconstructing the full variability of SO abyssal MOC using τ_x^{xavg} alone.

timescales shorter than approximately 2 years. When η serves as the sole predictor, DBNN accurately reconstructs the SO abyssal MOC with a median R^2 of 0.83. This result is consistent with the findings of Auger et al. (2024), which shows that bottom water production in the Ross Sea leaves fingerprints on η. The capability of η to reconstruct SO abyssal MOC variability is also in agreement with the zonal momentum balance mechanism discussed by Stewart and Hogg (2017) (see their Figure 1), which suggests the presence of a barotropic pressure signal along the AABW export pathways.

In summary, among all tested satellite-measurable quantities, p_b exhibits the greatest capability in reconstructing the variability of both the AMOC and SO MOCs. The combination of η and τ_x^{xavg} demonstrates comparable MOC reconstruction skill in the SO to that of p_b, suggesting that DBNN using η and τ_x^{xavg} may serve as a practical alternative given the uncertainty in p_b observation. Additionally, including θ_s and/or S_s yields no improvement in reconstruction skill, although in isolation they provide moderate predictive skill.

3.3. Significance of the Non-Locality of the DBNN

In this section, we explore whether and how the proposed non-local DBNN benefits from its ability to account for non-local information. To this end, we trained an additional set of DBNNs to reconstruct MOC variability at each latitude individually, using input features only from the same latitude. That is, instead of training a single DBNN

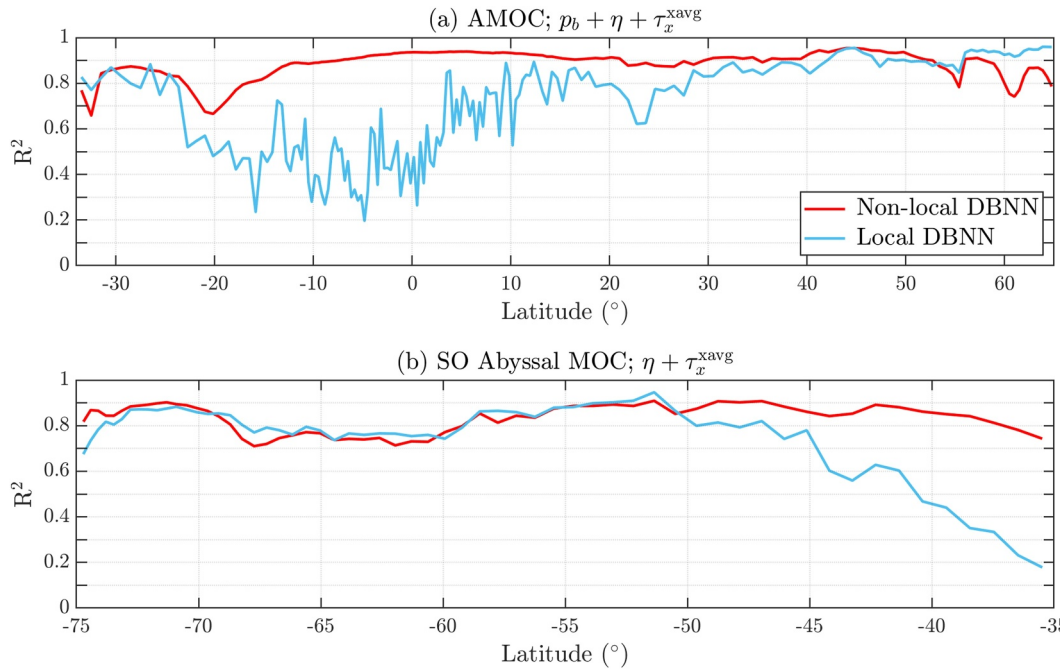

Figure 6. Comparison of the skill of the proposed non-local DBNN (red line) with a local version of the DBNN (cyan line) that reconstructs MOC viability at each latitude individually using input variables from the same latitude for the (a) AMOC and (b) SO abyssal MOC. The input variables used for AMOC reconstruction are p_b, η, and τ_x^{xavg}, while the input variables used for reconstructing SO abyssal MOC are η and τ_x^{xavg}. The 10-year LPF is applied to both the non-local and local DBNNs before the neural network training.

using input features from the entire basin to reconstruct MOC strength across n latitudes, we trained n separate DBNNs, each tasked with reconstructing MOC at a specific latitude using only input features from the same latitude. In Figure 6a, we compare the skill of this local version of the DBNN (cyan line) with the proposed non-local DBNN (red line) in reconstructing AMOC variability on decadal and longer timescales. Without the knowledge of non-local p_b, η, and τ_x^{xavg}, the reconstruction skill drops significantly near the equator for the long-term AMOC variability ($R^2 \sim 0.5$), presumably due to the breakdown of geostrophy, which is consistent with the findings in Hughes et al. (2018) and Solodoch et al. (2023). This indicates the importance of the non-locality in the proposed non-local DBNN, which is essential for its success in reconstructing equatorial AMOC variability with $R^2 \sim 0.9$.

We argue that the success of the non-local DBNN in the equatorial region relies on the coherence of long-term AMOC variability (Bingham et al., 2007; Gu et al., 2020; Wunsch & Heimbach, 2013). Or restated, AMOC variability at the equator can be predicted by the geostrophic balance at other latitudes where the variability is correlated with equatorial variability. Indeed, the squared cross-correlation of AMOC strength variability across latitudes (Figure 7c) shows that decadal and longer timescale AMOC variability at the equator is highly correlated with variability in the subtropical regions (e.g., the squared Pearson correlation coefficient $r^2 \sim 0.7$ between the equator and 26°N). The cross-correlation decreases at high latitudes in the Northern Hemisphere (e.g., $y > 55$°N), where the local DBNN outperforms the proposed non-local DBNN (Figure 6a). In this region, the inclusion of non-local information detracts from the training of the neural network, because much of the non-local information is irrelevant but significantly increases the size of the network. Furthermore, the coherence of AMOC variability diminishes on shorter timescales (Figures 7a and 7b). The lower cross-correlation of the full AMOC variability near the equator (e.g., $r^2 \sim 0.3$ between the equator and 26°N) explains the skill reduction of the DBNN when reconstructing full AMOC variability in the equatorial region (see the drop in the black line near the equator in Figure 4).

Another advantage of the non-locality is that it enables the DBNN to reconstruct the SO Abyssal MOC using η and τ_x^{xavg} without relying on p_b, satellite observations of which are subject to relatively large uncertainties (e.g., see the discussions in Jeffree et al., 2024). As shown in Figure 6b, the local DBNN with η and τ_x^{xavg} as input

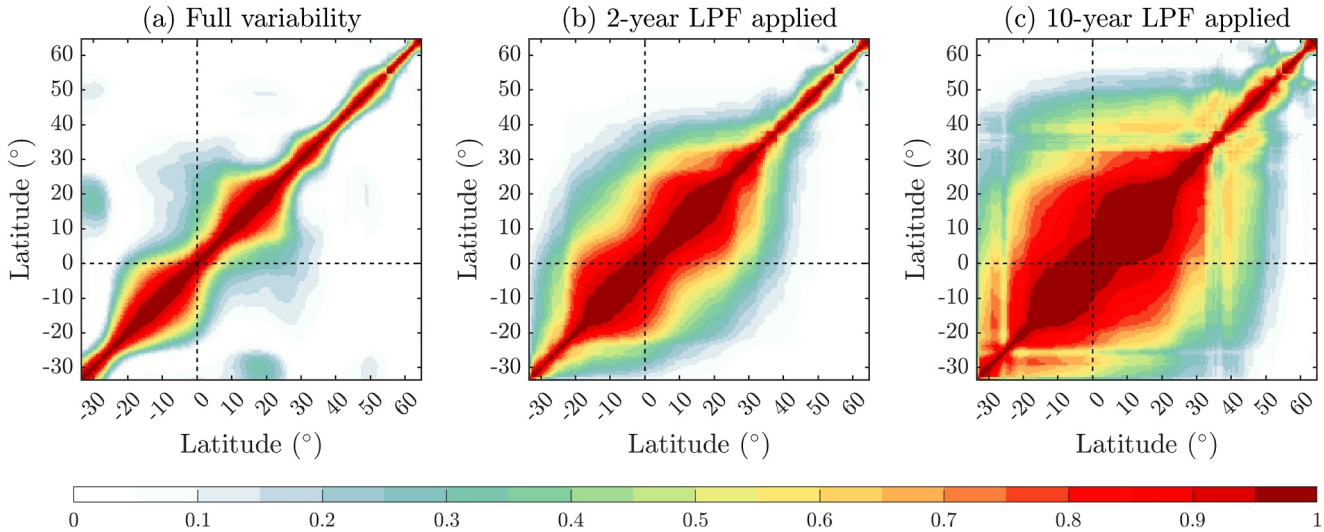

Figure 7. The cross-correlation of the detrended AMOC strength (a) without LPF applied, (b) with the 2-year LPF applied, and (c) with the 10-year LPF applied.

variables (cyan line) shows reduced capability in reconstructing SO Abyssal MOC north of 50°S, while the non-local DBNN (red line) maintains strong performance with $R^2 \sim 0.8$. The signal of AABW export in η is expected to be strongest where stratification is weakest and the flow is most barotropic (Stewart & Hogg, 2017). Consequently, the local η signals become less relevant for MOC variability toward lower latitudes, where stratification increases. This leads to the reduced skill of the local DBNN in reconstructing the SO Abyssal MOC to the north of ~50°S.

4. Interpreting the DBNN Using Layer-Wise Relevance Propagation

While we have demonstrated the promising capability of the proposed non-local DBNN in reconstructing MOC variability, like most data-driven approaches, the DBNN operates as a "black box," leaving the specific roles of each input feature (i.e., input variable from a specific location) unclear. In order to uncover which input features are most influential in driving the network's decisions and to gain a deeper understanding of underlying physical mechanisms learned by the DBNN, we apply a neural network interpretation technique, the Layer-wise Relevance Propagation (Bach et al., 2015), to the trained DBNN. The Layer-wise Relevance Propagation traces the contributions of each input feature by attributing the reconstructed MOC strength back to the input layer, generating a "relevance," $\mathcal{R}$, which quantifies how relevant each input feature is to the final MOC reconstruction. The relevance $\mathcal{R}$ shares the same units as the MOC strength, and its sum over all input features equals the predicted MOC strength, due to the conservation of relevance during attribution. A comprehensive review of the Layer-wise Relevance Propagation in the context of geophysical applications can be found in Section 3.2 of Toms et al. (2020). Since the DBNN has multiple output features (i.e., MOC strength time series at different latitudes), we apply the Layer-wise Relevance Propagation separately to each output, and we thus obtain the relevance maps for reconstructing MOC at each latitude. The Layer-wise Relevance Propagation is conducted on a sample-by-sample basis, meaning that $\mathcal{R}$ is a function of input features, output features, and time. We consider the time-mean absolute relevance $\overline{|\mathcal{R}|}^t$, which captures how input features influence the reconstructed MOC strength variability, irrespective of whether the MOC strength anomaly itself is positive or negative. The relevance shown in this section is computed from the baseline DBNN (i.e., with p_b, η, and τ_x^{xavg} as inputs).

4.1. The Relevance of Input Features in AMOC Reconstructions

Figure 8 presents the relevance maps of p_b in reconstructing AMOC variability. To generate a non-dimensional relevance maps, we first normalize the relevance by $\mathcal{R}_* = \overline{|\mathcal{R}|}^t / \sum \overline{|\mathcal{R}|}^t$, where $\sum$ indicates summation across all input features (i.e., $p_b(x, y)$, $\eta(x, y)$, and $\tau_x^{\text{xavg}}(y)$). We then compute the latitudinal mean of the normalized relevance, $\overline{\mathcal{R}_*}^y$, averaged within the latitude range indicated by the dashed lines. Larger values of $\overline{\mathcal{R}_*}^y$ indicate that the corresponding input feature plays a more significant role in MOC reconstruction within the dashed lines.

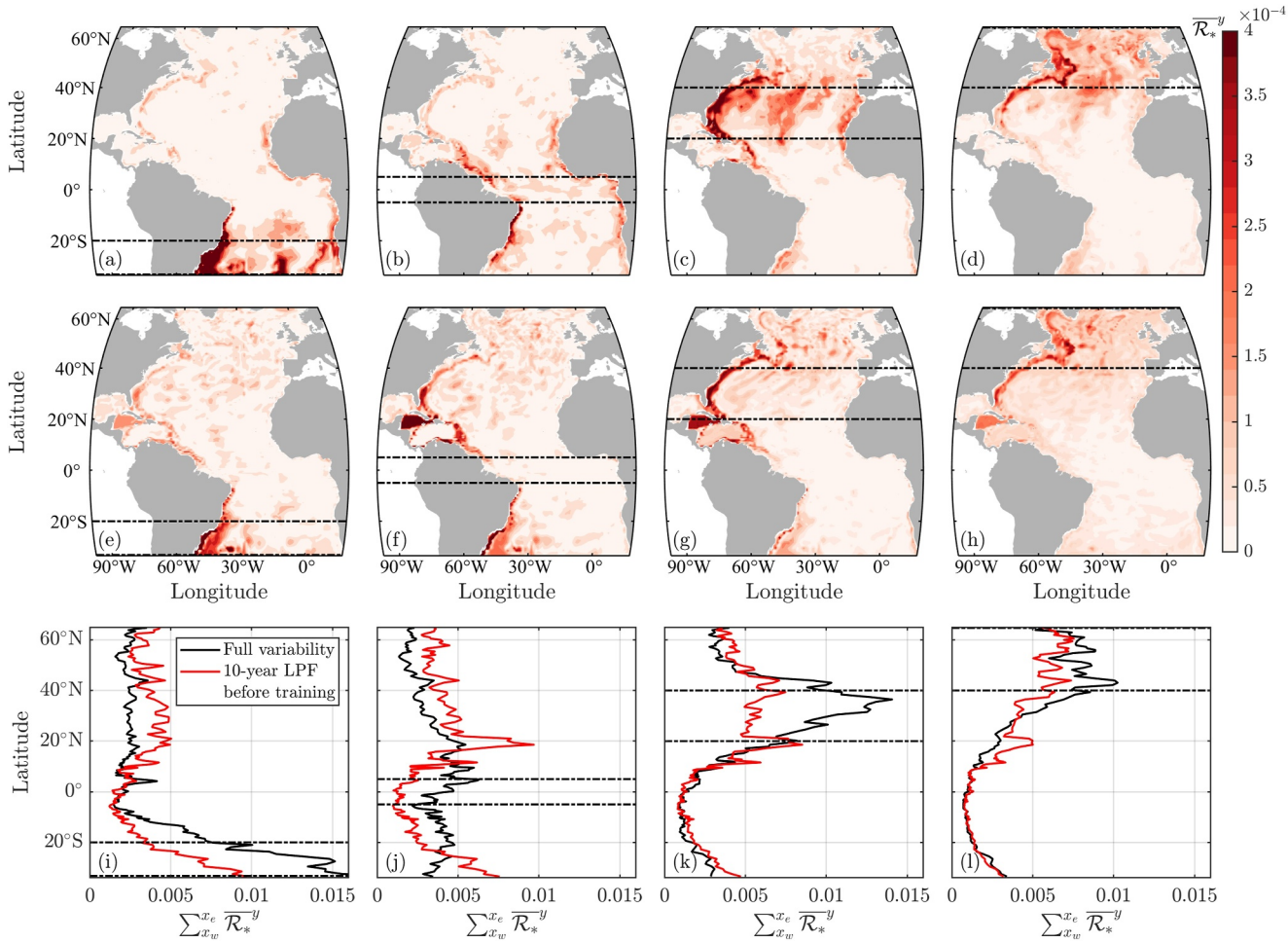

Figure 8. Normalized relevance of p_b for reconstructing the AMOC full variability averaged across various latitude ranges: (a) 34°S to 20°S, (b) 5°S to 5°N, (c) 20°N to 40°N, and (d) 40°N to 45°N, as indicated by the dashed lines. (e)–(h) As in (a)–(d), but with the 10-year LPF applied before DBNN training. The zonal summation of the relevance is shown in panels (i)–(l).

Note that the background noise in the relevance map may result from the imperfect MOC reconstruction by the DBNN; the primary focus is on the hotspots of the relevance.

In reconstructing the full AMOC variability in the subtropical region (Figures 8a and 8c), p_b near the western boundary shows strong relevance, while the eastern boundary and the mid-Atlantic ridge also exhibit higher relevance compared to flat, open ocean regions. This pattern aligns with the geostrophy argument for MOC reconstruction that the zonally integrated meridional transport at a given latitude and depth can be determined from the east-west bottom pressure difference across the basin, although the mid-Atlantic ridge complicates this argument since it can block specific depths at certain latitudes (Bingham & Hughes, 2009). For reconstructing AMOC strength variability on decadal and longer timescales (Figures 8e and 8g), the relevance at the mid-Atlantic ridge and eastern boundary diminishes, consistent with previous findings that p_b anomalies near the western boundary play a dominant role in controlling low-frequency AMOC variability (Bingham & Hughes, 2008, 2009), and that p_b anomalies on the eastern and western flanks of the mid-Atlantic ridge can be nearly equal, effectively canceling out in zonal pressure differential calculations (Bingham & Hughes, 2008; Landerer et al., 2015). The prominence of the western boundary is also supported by the understanding that it serves as the primary route for propagating disturbances triggered by deep water formation anomalies at high latitudes (Elipot et al., 2014), which propagate via coastal trapped waves and advection, as evidenced in both models and observations (Elipot et al., 2013; Johnson & Marshall, 2002; Kostov et al., 2023; Roussenov et al., 2008; van Sebille et al., 2011; Zhang, 2010). A significant area of high p_b relevance is observed in the northern Caribbean Sea (around 80°W, 18°N). However, this high relevance should be viewed with caution due to

previously documented inaccuracies in the Caribbean circulation within the low-resolution ACCESS Ocean Model 2 (Kiss et al., 2020). In the subpolar region (Figures 8d and 8h), the relevance of p_b is less concentrated near the western boundary, possibly reflecting the more complex interactions and coupling between the water mass transformation processes occurring in both the western and eastern subpolar gyres, which all contribute to the low-frequency AMOC variability (Kostov et al., 2023; Yeager et al., 2021).

When reconstructing the AMOC away from the equator (Figure 8, except for the second column), local information (i.e., relevance inside the dashed lines) is significantly more important than non-local information (i.e., relevance outside the dashed lines). However, when reconstructing equatorial AMOC variability (Figures 8b, 8f, and 8j), on the other hand, non-local p_b relevance overweights local relevance, which is consistent with the breakdown of local geostrophy and the failure of the local DBNN discussed in Section 3.3. This again suggests that the DBNN reconstructs equatorial AMOC variability via geostrophic balance at remote latitudes, which is possible given the coherence of AMOC variability (c.f. Figure 7). The zonal summation of p_b relevance (Figure 8j) indicates that non-local information becomes more relevant when constructing equatorial AMOC variability on decadal and longer timescale, in line with the fact that the AMOC is more latitudinally coherent on these time scales (c.f. Figure 7). Therefore, the Layer-wise Relevance Propagation further confirms the importance of non-locality in the proposed DBNN.

In Supporting Information S1 (Figures S10 and S11), we present relevance maps for η and τ_x^{xavg}. The magnitude of η relevance is notably lower than that of p_b, suggesting a relatively minor role of η in AMOC reconstruction. This is consistent with the importance of baroclinic processes in controlling the long-term variability of p_b and the AMOC reported in Bingham and Hughes (2009). Nevertheless, we acknowledge that η may become more relevant in high-resolution simulations (or the real oceans), where shelf processes and coastal sea level variability are better resolved (Higginson et al., 2015; Little et al., 2019). Additionally, the relevance of τ_x^{xavg} (Figure S10 in Supporting Information S1) shows a latitudinal pattern similar to the zonal summation of p_b relevance (Figures 8i–8l), though it diminishes significantly over longer timescales. In fact, the relevance of τ_x^{xavg} on decadal and longer timescales is more than an order of magnitude smaller than the zonal summation of p_b relevance. This weak relevance aligns with the limited capability of τ_x^{xavg} in capturing the low frequency variability of the AMOC, the minor improvement when it serves as an additional input variable alongside p_b (c.f. Figure 5c), and the RAPID array observations that wind forcing primarily drives the short term AMOC variability at 26.5°N (Frajka-Williams et al., 2019; Pillar et al., 2016). Note that the comparison between the magnitudes of τ_x^{xavg} and the zonal summation of p_b relevance should be viewed with caution, as the input size for p_b and τ_x^{xavg} differs (i.e., p_b has zonal dependence), and the regularization applied during neural network training prevents the DBNN from being dominated by any single input variable (i.e., if duplicated τ_x^{xavg} were added to the inputs, the total relevance of τ_x^{xavg} would be expected to increase).

In summary, via Layer-wise Relevance Propagation, we find that AMOC reconstruction by the DBNN is largely dominated by p_b signals from the western boundary, particularly for long-term variability. While the DBNN primarily relies on local p_b signals to reconstruct the AMOC outside the equatorial region, non-local information becomes more significant than local information for reconstructing AMOC variability near the equator.

4.2. The Relevance of Input Features in SO MOC Reconstructions

In contrast to the relatively minor role of η in AMOC reconstruction, η shows high relevance in reconstructing the full SO abyssal MOC variability at high latitudes from 75°S to 65°S, even outweighing the relevance of p_b (Figures 9a and 9e). Elevated η relevance is observed in the Weddell Sea (around 50°W) and the Ross Sea (around 180°), which aligns with the primary sites of dense water formation and export (Purkey et al., 2018). However, as the 1° grid spacing of the ACCESS-ESM1.5 ocean model cannot fully resolve the surface water mass transformation occurring on the Antarctic continental shelf (Kiss et al., 2020), the elevated η relevance in the Weddell and the Ross Seas may instead be related to an accumulation of baroclinic coastal trapped waves that propagate around Antarctica. The relevance of η reduces in the northern portion of the SO abyssal MOC while p_b relevance increases, consistent with the failure of local DBNNs using η and τ_x^{xavg} to reconstruct the SO abyssal MOC from 45°S to 35°S (see the cyan line in Figure 6b).

The relevance of p_b in reconstructing the northern portion of the SO abyssal MOC exhibits clear zonal dependence (Figures 9c and 9d). Regions with high p_b relevance align with the primary AABW export routes identified by

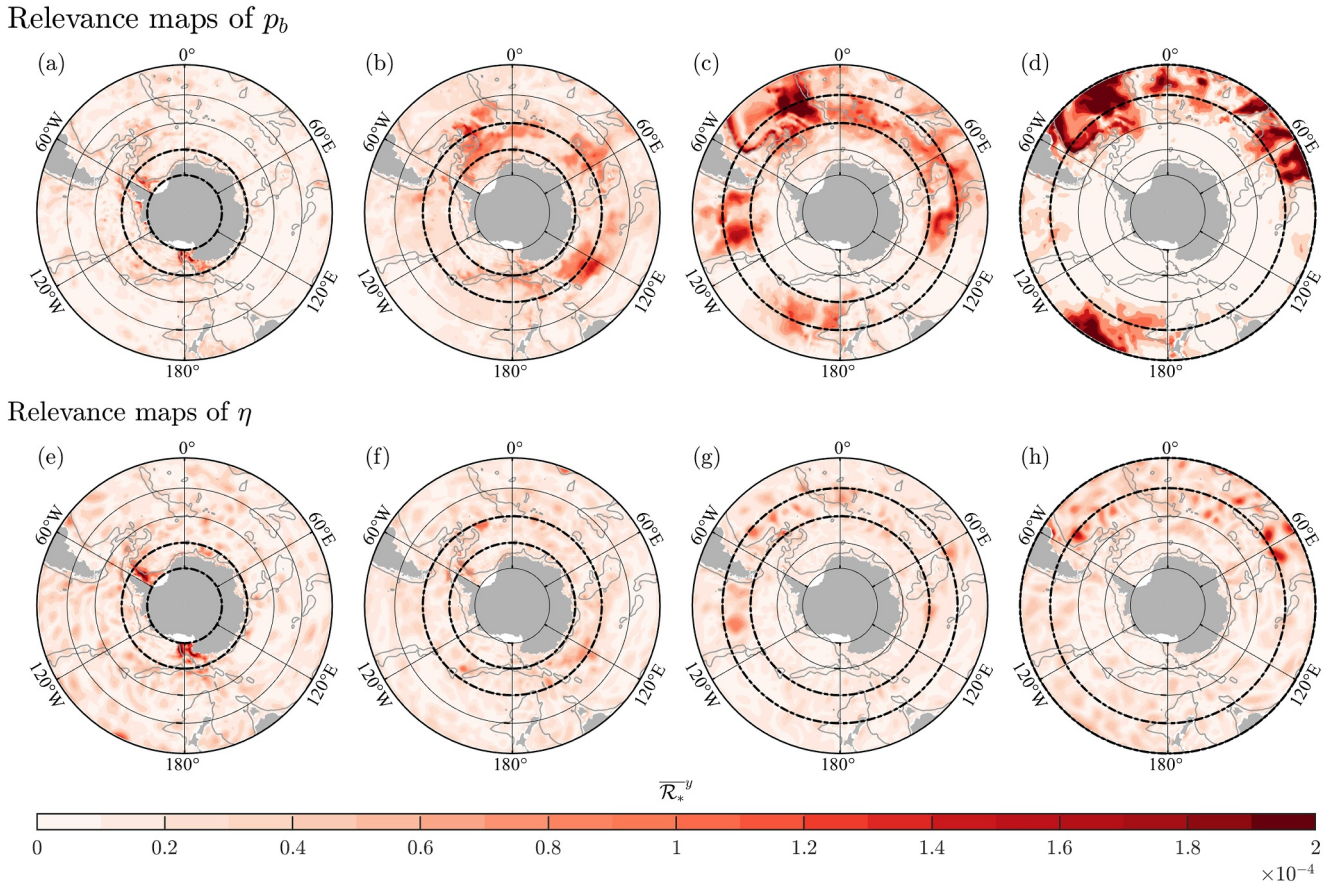

Figure 9. Normalized relevance of p_b for reconstructing SO abyssal MOC full variability averaged across various latitude ranges: (a) 75°S to 65°S, (b) 65°S to 55°S, (c) 55°S to 45°S, and (d) 45°S to 35°S, as indicated by the dashed lines. (e)–(h) As in (a)–(d), but show the relevance of η. Note that the background noise in the relevance map may result from the imperfect MOC reconstruction by the DBNN; the primary focus is on the hotspots of the relevance. Gray contours indicate the 3-km isobath.

Solodoch et al. (2022), such as the broad belts west of the Mid-Atlantic Ridge in the Atlantic Ocean (approximately 50°W to 10°W) and east of the Campbell Plateau in the Pacific Ocean (approximately 180°W to 150°W). This may suggest that the SO abyssal MOC can be effectively reconstructed via geostrophic balance along these primary export routes, which are guided by prominent topographic features.

In reconstructing the SO mid-depth MOC, p_b dominates the relevance, and shows hotspots near the coast, such as around Cape Horn and New Zealand (Figure S12 in Supporting Information S1). This indicates that the SO mid-depth MOC can be reconstructed via the p_b difference between the "eastern" and "western" boundaries of the basin, similar to the findings in the AMOC reconstruction.

5. Reconstructing Density-Dependent MOC Variability

In previous sections, we explored the reconstruction of MOC strength across various latitudes and basins, demonstrating the promising capabilities of the proposed non-local DBNN. Our results revealed that geostrophic balance is the dominant physical mechanism driving the DBNN's predictions. We also showed that non-locality of the DBNN, combined with the coherence of MOCs, is essential for accurately reconstructing equatorial AMOC strength, where the local geostrophic balance breaks down.

However, the vertical structure of the MOC may change due to both intrinsic and forced variability, particularly under climate change scenarios. Consequently, simplifying MOC variability to changes in its "strength" does not fully capture its structural changes and the associated impacts on tracer transport. We now explore whether the data-driven approach can be applied to reconstruct the density-dependent MOC variability (rather than just MOC strength variability) within the density range of interest. A straightforward extension of the DBNN introduced in

Section 2.3 to reconstruct the density-dependent MOC can be achieved by simply adjusting the output via replacing MOC strength $\psi(y, t)$ with the depth-dependent streamfunction $\Psi(y, \sigma_2, t)$ as defined in Section 2.2. In this case, the output size of the DBNN would scale from the number of latitudes to the product of the numbers of latitudes and density levels (i.e., each output neuron represents the MOC variability at a specific latitude and density level). The MOC reconstruction skill of this straightforward extension of the baseline DBNN is shown in Figure S13 in Supporting Information S1.

Nevertheless, we find that a slight modification to the DBNN pipeline helps to improve its skill in reconstructing density-dependent MOC compared to this straightforward extension (R^2 increases by ~0.1). Specifically, instead of directly reconstructing MOC variability at each latitude and density level, we propose to reconstruct the time-varying strength of the empirical orthogonal functions (EOFs) of MOC variability $\Psi(y, \sigma_2, t)$ (Weare & Nasstrom, 1982). We first derive EOFs from the training data set that capture 95% of the MOC variability (around 30–60 EOFs depending on the MOC domain and timescale). Increasing the number of EOFs to capture 99% of the MOC variability (around 100–200 EOFs) yields nearly identical reconstruction skill (see Figure S14 in Supporting Information S1). The DBNN is then trained to predict the temporal evolution of the magnitudes of these EOFs. For MOC reconstruction on the testing data set, the trained DBNN predicts the time-varying EOF magnitudes, which are combined with the EOFs derived from the training data set to reconstruct the density-dependent $\Psi(y, \sigma_2, t)$. The use of EOFs can be thought of as a linear dimensionality reduction technique. As the EOFs form a complete basis for the MOC latitude/density structure, they can, in principle, be used to reconstruct the MOC in any model or in nature. However, this reconstruction may require more EOFs to achieve the same accuracy when the EOFs from one model are used to reconstruct the MOC in another. Additionally, based on hyperparameter tuning, we removed the hidden layer with 64 neurons in the non-linear branch. The input variables remain the same as the baseline DBNN introduced in Section 2.3, that is, p_b, η, and τ_x^{xavg} from the corresponding MOC domain (non-local information included). The density ranges considered are from 35 to 36.9 kg/m^3 for the AMOC, and from 35 to 37.2 kg/m^3 for the SO MOCs. We refer to this slightly modified neural network as EOF-DBNN.

In Figure 10, we show the skill of EOF-DBNN in reconstructing the density-dependent variability of both the SO MOCs and the AMOC. Animations comparing diagnosed and reconstructed MOC on decadal and longer timescales during the last 200 model years can be found in Supporting Information S1 (Movies S1 and S2). Overall, the EOF-DBNN captures a large portion of the MOC variability across different timescales, except in regions with very weak mean streamfunction (e.g., $|\overline{\Psi}^t| < 2$ Sv; hatched region). The density-dependent reconstruction skill of the EOF-DBNN shows an analogous pattern to the performance of the baseline DBNN in reconstructing the MOC strength (Figure 4). For example, the reconstruction skill by the EOF-DBNN tends to decrease toward longer timescales in the Southern Ocean, and a notable skill drop is observed in reconstructing the SO abyssal MOC South of 65° (c.f. Section 3.1). A skill reduction is also found near the equator across all density levels when reconstructing the full AMOC variability, and this skill drop disappears when focusing on variability over inter-annual or longer timescales. This can again be attributed to the weaker coherence of AMOC variability in shorter timescales, which reduces the efficacy of non-local p_b information (c.f. Section 3.3). For decadal and longer-term variability, the AMOC at denser density levels (e.g., $\sigma_2 > 36$ kg/m^3) is better constrained by the EOF-DBNN compared to lighter density levels, which is likely due to p_b providing more geostrophic flow information at denser layers. Additionally, the relationship between p_b and MOC variability in the deep ocean may be less complex due to the reduced density variations compared to the upper ocean, making it easier for the EOF-DBNN to capture the relationship. Additionally, we note that the reconstruction by the EOF-DBNN tends to slightly underestimate the variance of the diagnosed MOC (Figure S16 in Supporting Information S1), similar to the underestimated high-amplitude events in the MOC strength reconstruction (Figure 3). Additional tests (not shown) suggest that this underestimation may be partly attributed to the regularization techniques applied during DBNN training (c.f. Section 2.3.2), though the exact causes remain unclear.

An additional test using only p_b as the input variable for the EOF-DBNN shows nearly identical reconstruction skill (Figure S15 in Supporting Information S1), once again underscoring the critical role of p_b in MOC reconstruction. This capability aligns with the notion that p_b signals from different depths over the continental slope can be combined to estimate ocean transport at various depths (Hughes et al., 2018), though the definition of MOC in density coordinates makes this interpretation less straightforward.

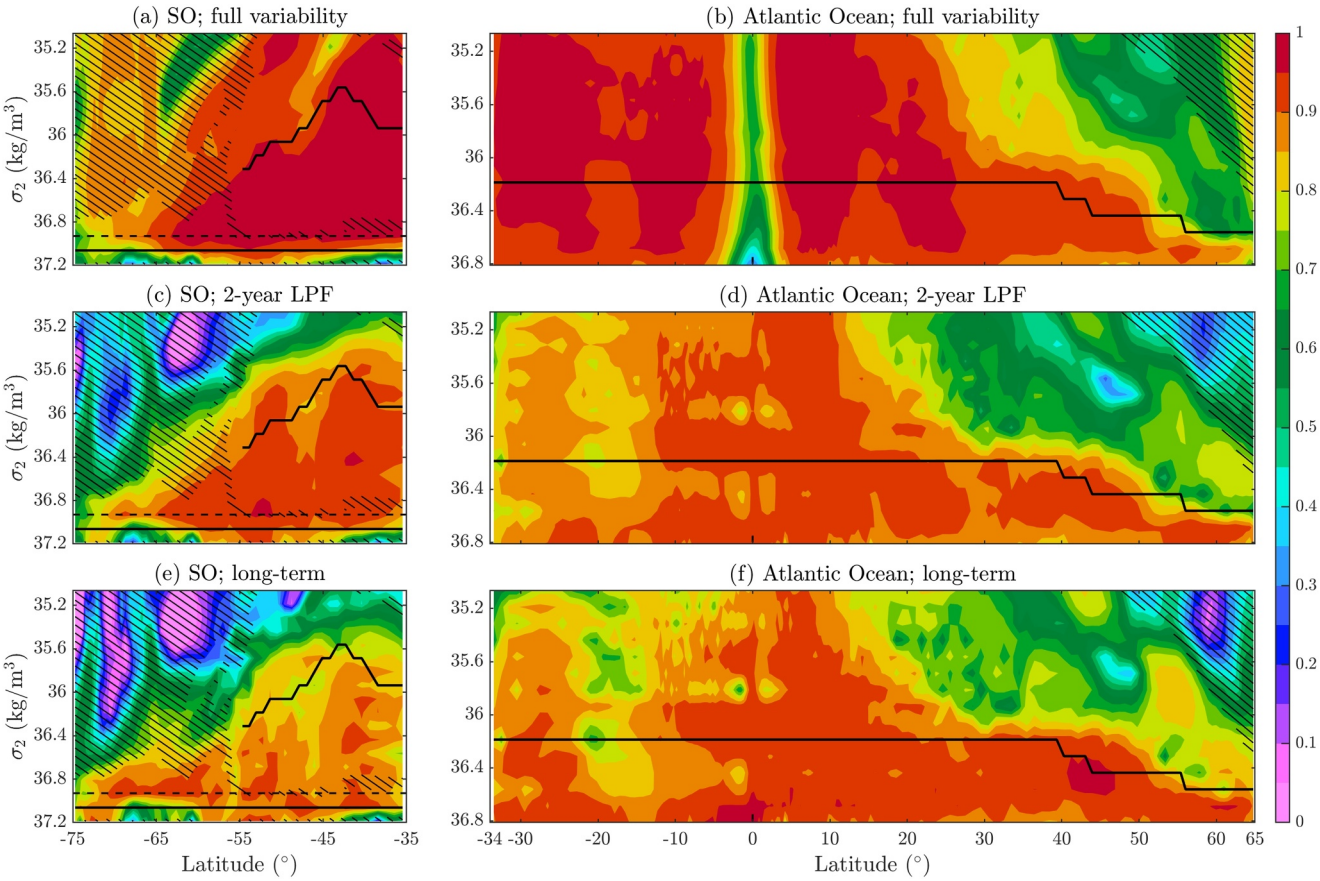

Figure 10. Reconstruction skill (R^2) of the EOF-DBNN detailed in Section 5 for reconstructing the density-dependent full variability of the (a) SO MOCs and (b) AMOC in the ACCESS-ESM1.5 PI control simulation. (c)–(d) As in (a)–(b), but for variability on inter-annual and longer timescales. (e)–(f) As in (a)–(b), but for variability on decadal and longer timescales. Regions with time-mean streamfunction less than 2 Sv are hatched. The black line represents the density level at which the MOC strength is defined (i.e., the same as the green line in Figure 1).

6. Reconstructing Density-Dependent MOC Variability in GFDL-CM4

The preceding sections have shown that data-driven approaches can effectively reconstruct both MOC strength and density-dependent MOC variability, based on the PI control simulation from ACCESS-ESM1.5. However, since model biases are inevitable, it is important to validate the proposed approach using additional climate models. In this section, we explore whether the proposed data-driven approach remains effective using the PI control simulation from GFDL-CM4 (see Section 2.1 for model details). The time-mean MOCs in the Atlantic and the Southern Ocean from GFDL-CM4 are presented in Supporting Information S1 (Figure S1). The density levels corresponding to the maxima of the mid-depth cells are denser in GFDL-CM4 compared to ACCESS-ESM1.5, and the SO mid-depth MOC is notably weaker in GFDL-CM4. The weaker SO mid-depth MOC may stem from the better-resolved mesoscale eddies in GFDL-CM4. However, a comprehensive analysis of the underlying reasons would require further investigation into model configurations, parameterizations, and physical processes, which is outside the immediate focus of this study.

Figure 11 shows the skill of the EOF-DBNN in reconstructing the density-dependent MOC variability in both the Southern Ocean and the Atlantic Ocean, with the EOF-DBNN re-trained using data from the GFDL-CM4 PI control run. The density range in the SO is refined to [35, 37.1] kg/m^3, as the time-mean streamfunction at densities higher than 37.1 kg/m^3 can reach zero north of 56°S in the GFDL-CM4 PI control run. The AMOC reconstruction skill is very similar to that in the ACCESS-ESM1.5 (Figure 10), though there is a slight decline in skill at high latitudes (e.g., to the north of 55°N) and an improvement in the Southern Hemisphere, particularly on longer timescales. For SO mid-depth MOC reconstruction, the skill is comparable to that in ACCESS-ESM1.5, but generally lower. For instance, the R^2 is approximately 0.7 for reconstructing inter-annual and longer-term

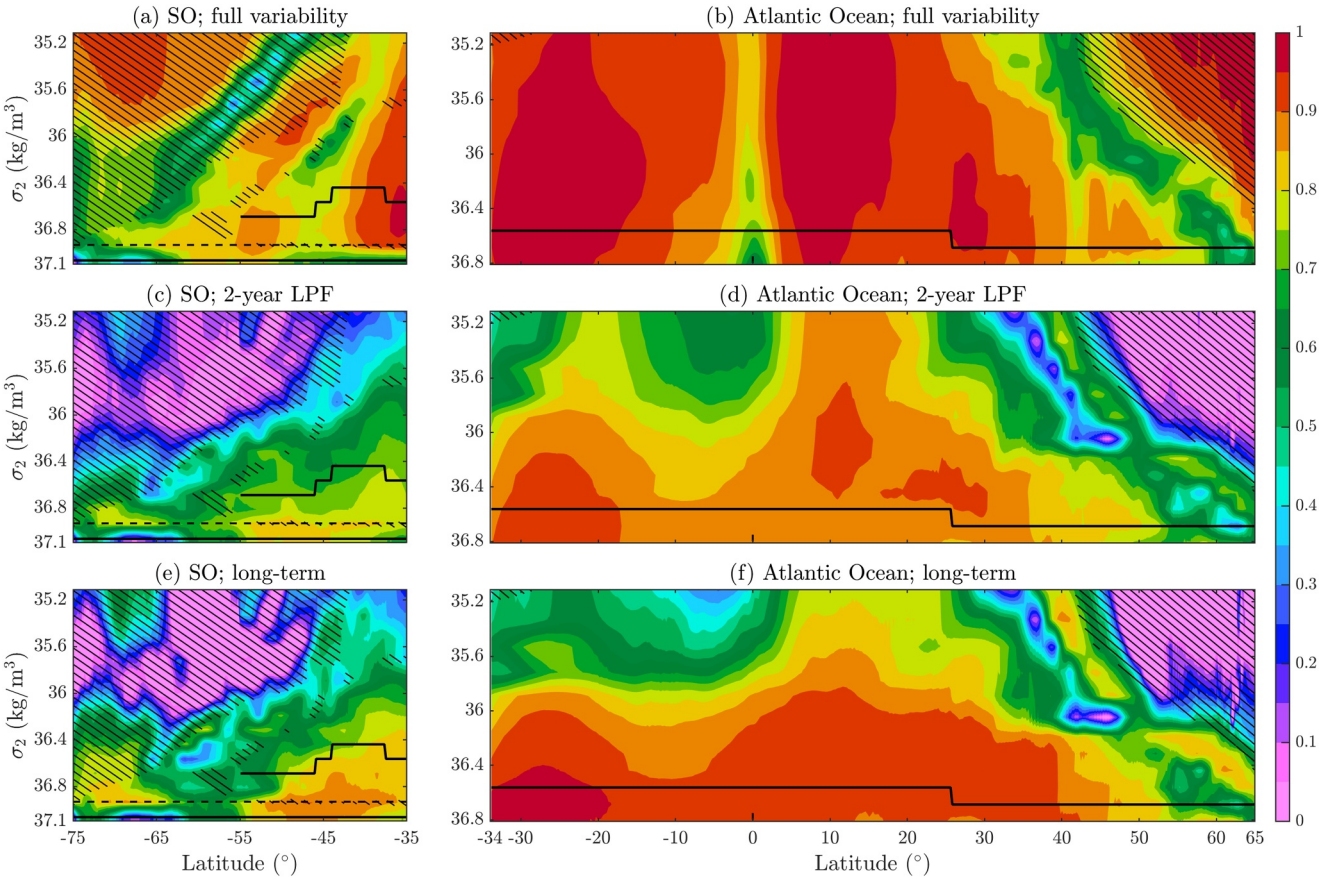

Figure 11. As in Figure 10, but for the GFDL-CM4 PI control simulation.

variability of the SO mid-depth MOC in GFDL-CM4, whereas it reaches 0.9 in ACCESS-ESM1.5. This potentially suggests that the proposed DBNN may be less efficient when mesoscale eddies are partially resolved, and convolutional neural networks might be more suitable (Meng et al., 2024).

A substantially lower skill is observed when reconstructing the SO abyssal MOC at high latitudes (e.g., around 65°S). While a similar skill drop at high latitudes was also observed in ACCESS-ESM1.5 (Figure 10), it is much more pronounced in GFDL-CM4 (Figure 11). The issue of variance underestimation is also more evident in GFDL-CM4 for the SO abyssal MOC reconstruction (Figure S17 in Supporting Information S1). We note that the diagnosed SO abyssal MOC anomalies in GFDL-CM4 exhibit distinct variability characteristics across each 100-year subset (Figure S18a in Supporting Information S1). This suggests that the assumption of identical distributions across subsets inherent to cross-validation (c.f. Section 2.3.2) does not hold. Consequently, the neural network training during cross-validation suffers from an out-of-distribution issue, where the test data set includes previously unseen conditions outside the range of the training data. To confirm that the low reconstruction skill is indeed caused by the out-of-distribution issue, we performed an additional test by randomly shuffling the 500-year data set and allocating two-thirds for training and one-third for testing. This approach leads to much better reconstruction of the MOC variability (Figures S18c and S18e in Supporting Information S1). However, we emphasize that this artificial manipulation ensures identical training and testing distributions but introduces data leakage, making it impractical for real-world MOC reconstruction.

7. Summary and Discussion

The meridional overturning circulation (MOC) plays a crucial role in regulating the Earth's climate, but direct observations of the MOC remain sparse and geographically confined. In this study, we developed a data-driven approach to indirectly monitor MOC variability using satellite-measurable quantities, based on two preindustrial control simulations from CMIP6. We demonstrated that a spatially non-local dual-branch neural network

(DBNN; Figure 2) can accurately reconstruct MOC strength variability across various latitudes, timescales, and ocean basins (Figure 4) using ocean bottom pressure, sea surface height, and zonally-averaged zonal wind stress as input variables. We further show that the DBNN can be extended to reconstruct MOC variability across different density levels, leading to a density-dependent MOC reconstruction (Figures 10 and 11).

Our results highlight the critical role of ocean bottom pressure, particularly near the western boundary (see also Bingham & Hughes, 2008) and dense water export pathways, in reconstructing MOC variability (Figures 5, 8, and 9). The geostrophic balance was identified as the dominant physical mechanism driving the DBNN's predictions (Figures 8 and 9). The non-locality of the DBNN and the coherence of AMOC variability (Figure 7) are essential for accurately reconstructing MOC strength near the equator, where local geostrophic balance breaks down (Figure 6). Given the relatively large uncertainties in ocean bottom pressure observations from satellites (Jeffree et al., 2024), combinations of input variables excluding ocean bottom pressure may offer alternative solutions. While the DBNN using only sea surface height and zonal wind stress accurately reconstructs MOCs in the Southern Ocean (Figures 5a and 5b), ocean bottom pressure remains indispensable in AMOC variability reconstruction (Figure 5c). This highlights the significant value of GRACE satellite data and emphasizes the need to improve the accuracy of these measurements.

Despite these achievements, significant challenges remain in realizing a comprehensive, satellite-based system for monitoring global MOC variability. This study relies on numerical simulations to demonstrate the strong relationship between satellite-measurable quantities and MOC strength anomalies on decadal and longer timescales. However, how to practically apply the proposed data-driven approach to real-world MOC monitoring remains an open question. A feasible pathway would likely still depend on numerical models, with real-world MOC observations (c.f. Section 1), serving as a valuable, independent, validation and testing data set (rather than training data). While numerical models may not capture the exact MOC variability observed in the real ocean, they are expected to correctly represent the leading-order physical processes governing MOC changes. Our results indicate that the neural network might effectively learn these physical relationships, suggesting the potential for applying a model-trained network to reconstruct real-world MOC by leveraging both learned relationships and satellite observations. Nonetheless, future validation using observations is essential, as the neural networks trained in models may also learn non-physical relationships arising from intrinsic model biases (Jackson et al., 2023), such as the spurious numerical mixing in overflows in z-coordinate models (Wang et al., 2015).

A key step toward real-world application is the development of neural networks that are not model-specific. To primarily explore the possibility, we interpolated data from GFDL-CM4 and ACCESS-ESM1.5 onto the same 1° grid and trained the DBNN separately for each model to reconstruct AMOC strength variability. We then tested the capability of a network trained on one model to reconstruct AMOC variability in the other model, without prior exposure to its data. Although cross-model reconstruction skill was lower than same-model reconstruction (Figures S19 in Supporting Information S1), the DBNN trained on ACCESS-ESM1.5 still achieved $R^2 \sim 0.6$ across most latitudes when applied to GFDL-CM4. Reconstruction skill was lower in the reverse case, particularly for long-term variability, but remained non-negligible. These preliminary results suggest the possibility of developing model-independent neural networks for MOC reconstruction, but further efforts are needed to incorporate a broader range of climate models, refine input feature engineering, and potentially leverage transfer learning. Furthermore, adapting the approach to accommodate satellite-measurable quantities at realistic spatial and temporal resolutions is essential for real-world applications.

Moreover, while the EOF-DBNN demonstrates strong performance in reconstructing the full-depth MOC variability within individual models, its generalization capability warrants further evaluation. As EOFs derived from one model typically capture less variance when applied to other models or observational data, this limitation may hinder the DBNN's generalization capability. Future research could address this by training the network on EOFs derived from multiple models or by exploring alternative dimensionality reduction techniques, such as autoencoders.

Another major challenge is assessing whether the strong relationship between satellite-measurable quantities and MOC variability persists in a changing climate. The MOC is projected to undergo significant changes in structure and magnitude under a warming climate. For example, Antarctic Bottom Water has already shown signs of warming, freshening, and thinning in recent decades (Purkey et al., 2018). Although our density-dependent MOC reconstructions offer a promising approach to capture these changes, shifts in ocean density stratification could alter the relationship between satellite-measurable quantities (e.g., ocean bottom pressure) and MOC variability,

reducing the effectiveness of the proposed data-driven approach. One possible way to mitigate this issue is to leverage ongoing hydrographic sampling activities, such as the deep hydrographic sampling performed repeatedly along WOCE lines. Future research should explore how incorporating spatio-temporally sparse hydrographic data as additional inputs to the neural network might enhance its predictive skill, particularly under climate change scenarios like those described by the Shared Socioeconomic Pathways (Riahi et al., 2017). The challenge related to the out-of-distribution problem, as discussed in Section 6, could also potentially be alleviated by the inclusion of deep-ocean stratification observations. Additionally, while detrending was used in this study to remove model drift in preindustrial control simulations, alternative approaches, such as subtracting basin-wide or global means, may be more appropriate in climate change scenarios, where the trends are dynamically meaningful.

Moreover, our study focused primarily on the Atlantic and the Southern Ocean, leaving the Indo-Pacific MOC reconstruction yet to be investigated. The Indo-Pacific MOC has been shown to respond to AMOC changes on decadal to centennial timescales, effectively redistributing heat between the Atlantic and the Indo-Pacific basins (Sun & Thompson, 2020; Sun et al., 2022), and affecting the marine carbon cycle (Sun et al., 2024). The importance of this interbasin MOC response highlights the need to achieve a full global MOC monitoring system.

Additionally, we observed degraded skill in reconstructing Southern Ocean MOCs in a mesoscale-permitting climate model, relative to the baseline, non-eddying model studied here. Future work could explore whether adopting the convolutional neural network approach would mitigate this deficiency, as Meng et al. (2024) shows that convolutional neural networks better reconstruct the Southern Ocean MOCs compared with a fully-connected feed-forward neural network in their idealized eddy-resolving simulations. Additionally, the results from Meng et al. (2024) indicate that both ocean bottom pressure and sea surface height are essential for reconstructing Southern Ocean MOCs, while our results show that the DBNN, using ocean bottom pressure alone as the predictor, captures approximately 80% of the MOC variability, with the inclusion of sea surface height only marginally improving the reconstruction (Figure 5). These seemingly divergent results can be attributed to several potential reasons. First, the realistic bathymetry of the climate simulation used in this study allows ocean bottom pressure to encompass a broader range of density and depth levels, potentially providing more information about the geostrophic flow. Second, the 6-km resolution employed in Meng et al. (2024) resolves mesoscale eddies, unlike the non-eddying and eddy-permitting climate simulations used in our study (c.f. Section 2.1). Further investigation is needed to understand whether, and to what extent, a rich eddy field impacts MOC reconstruction in realistic simulations. Lastly, differences in neural network architectures might also contribute to these contrasting results.

Finally, the current input variables in the DBNN are in-phase with MOC predictions, yet a number of studies indicate that MOC responses can be preceded by changes in ocean properties, such as buoyancy anomalies at high latitudes (e.g., Kostov et al., 2023; Yeager et al., 2021). We attempted to augment the baseline DBNN by including sea surface temperature and salinity at one or several lagged times ranging from 6 months to 10 years as additional input variables, but this did not lead to enhanced reconstruction skill. We also attempted to include lagged sea surface temperature and salinity in the DBNN that originally used in-phase temperature and salinity as inputs, which did improve reconstruction skill, particularly at low latitudes, though the skill remained much lower than that of the proposed baseline DBNN. Future work should explore whether recurrent neural networks (e.g., Long Short-Term Memory neural networks; Hochreiter & Schmidhuber, 1997) could achieve further improvements compared to the baseline DBNN in this study, especially under climate change scenarios.

Appendix A: Formulation of the DBNN

Precisely, the DBNN introduced in Section 2.3 can be mathematically expressed as

First Hidden Layer:

$$z_i^1 = \sum_j W_{ij}^0 X_j + b_i^0, \quad i = 1, \ldots, 64, \tag{A1a}$$

Linear Branch:

$$z_i^{\text{linear}} = \sum_j W_{ij}^{\text{linear}} z_j^1 + b_i^{\text{linear}}, \quad i = 1, \ldots, n, \tag{A1b}$$

Non-linear Branch, Activation:

$$a_i^{\text{NL1}} = \sigma\left(z_i^1\right), \quad i = 1, \ldots, 64, \tag{A1c}$$

Non-linear Branch, Layer 1:

$$z_i^{\text{NL1}} = \sum_j W_{ij}^{\text{NL1}} a_j^{\text{NL1}} + b_i^{\text{NL1}}, \quad i = 1, \ldots, 64, \tag{A1d}$$

Non-linear Branch, Activation:

$$a_i^{\text{NL2}} = \sigma\left(z_i^{\text{NL1}}\right), \quad i = 1, \ldots, 64, \tag{A1e}$$

Non-linear Branch, Layer 2:

$$z_i^{\text{NL2}} = \sum_j W_{ij}^{\text{NL2}} a_j^{\text{NL2}} + b_i^{\text{NL2}}, \quad i = 1, \ldots, n, \tag{A1f}$$

Final Output:

$$Y_i = z_i^{\text{NL2}} + z_i^{\text{linear}}, \quad i = 1, \ldots, n. \tag{A1g}$$

Here X_j represents the one-dimensional input array, W_{ij} is the weight parameter, b_i is the bias parameter, superscripts indicate different neural network layers, n denotes the output dimension, and Y_i is the final output of the neural network.

Data Availability Statement

The ACCESS-ESM1.5 and GFDL-CM4 preindustrial control runs used in this study are publicly available from the CMIP6 website https://aims2.llnl.gov/search/cmip6/. The neural network and training pipeline are implemented using the Keras API. The empirical orthogonal functions used in Section 5 are computed using the scikit-learn package (Pedregosa et al., 2011), and the Layer-wise Relevance Propagation is implemented via the innvestigate package (Alber et al., 2019) within Python. The Python scripts that were used to train the neural networks in this study have been published at Wei (2024). The data and MATLAB scripts used to generate the figures in the manuscript have been published at Wei (2025).

Acknowledgments

The authors thank Rory Bingham, Andrew Styles, and an anonymous reviewer for their thoughtful reviews and constructive comments, which improved the manuscript. This work used Bridges-2 at Pittsburgh Supercomputing Center (Brown et al., 2021) through allocation EES240055 from the Advanced Cyberinfrastructure Coordination Ecosystem: Services & Support (ACCESS) program (Boerner et al., 2023), which is supported by National Science Foundation Grants #2138259, #2138286, #2138307, #2137603, and #2138296. HW and ALS are partially supported by the National Aeronautics and Space Administration ROSES Physical Oceanography program under Grant 80NSSC23K0357. KS acknowledges support from the Office of Naval Research (ONR) Multidisciplinary University Research Initiative (MURI) Grant N00014-20-1-2023.

References

Abernathey, R., Marshall, J., & Ferreira, D. (2011). The dependence of Southern Ocean meridional overturning on wind stress. *Journal of Physical Oceanography*, *41*(12), 2261–2278. https://doi.org/10.1175/JPO-D-11-023.1

Adcroft, A., Anderson, W., Balaji, V., Blanton, C., Bushuk, M., Dufour, C. O., et al. (2019). The GFDL global ocean and sea ice model OM4.0: Model description and simulation features. *Journal of Advances in Modeling Earth Systems*, *11*(10), 3167–3211. https://doi.org/10.1029/2019MS001726

Alber, M., Lapuschkin, S., Seegerer, P., Hägele, M., Schütt, K. T., Montavon, G., et al. (2019). iNNvestigate neural networks. *Journal of Machine Learning Research*, *20*(93), 1–8. https://doi.org/10.5555/3327345.3327503

Ansorge, I. J., Baringer, M. O., Campos, E. J. D., Dong, S., Fine, R. A., Garzoli, S. L., et al. (2014). Basin-wide oceanographic array bridges the South Atlantic. *Eos, Transactions American Geophysical Union*, *95*(6), 53–54. https://doi.org/10.1002/2014EO060001

Auger, M., Spence, P., & Morrison, A. K. (2024). Surface signature of Dense Shelf Water overflows from satellite altimetry. In *2024 ocean sciences meeting*.

Bach, S., Binder, A., Montavon, G., Klauschen, F., Müller, K.-R., & Samek, W. (2015). On pixel-wise explanations for non-linear classifier decisions by layer-wise relevance propagation. *PLoS One*, *10*(7), e0130140. https://doi.org/10.1371/journal.pone.0130140

Bennetts, L. G., Shakespeare, C. J., Vreugdenhil, C. A., Foppert, A., Gayen, B., Meyer, A. K., et al. (2024). Closing the loops on Southern Ocean dynamics: From the circumpolar current to ice shelves and from bottom mixing to surface waves. *Reviews of Geophysics*, *62*(3), e2022RG000781. https://doi.org/10.1029/2022RG000781

Bentel, K., Landerer, F. W., & Boening, C. (2015). Monitoring Atlantic overturning circulation and transport variability with GRACE-type ocean bottom pressure observations–a sensitivity study. *Ocean Science*, *11*(6), 953–963. https://doi.org/10.5194/os-11-953-2015

Bingham, R. J., & Hughes, C. W. (2008). Determining North Atlantic meridional transport variability from pressure on the western boundary: A model investigation. *Journal of Geophysical Research*, *113*(C9), C09008. https://doi.org/10.1029/2007JC004679

Bingham, R. J., & Hughes, C. W. (2009). Geostrophic dynamics of meridional transport variability in the subpolar North Atlantic. *Journal of Geophysical Research*, *114*(C12), C12029. https://doi.org/10.1029/2009JC005492

Bingham, R. J., Hughes, C. W., Roussenov, V., & Williams, R. G. (2007). Meridional coherence of the north Atlantic meridional overturning circulation. *Geophysical Research Letters*, *34*(23), L23606. https://doi.org/10.1029/2007GL031731

Bishop, S. P., Gent, P. R., Bryan, F. O., Thompson, A. F., Long, M. C., & Abernathey, R. (2016). Southern Ocean overturning compensation in an eddy-resolving climate simulation. *Journal of Physical Oceanography*, *46*(5), 1575–1592. https://doi.org/10.1175/JPO-D-15-0177.1

Boerner, T. J., Deems, S., Furlani, T. R., Knuth, S. L., & Towns, J. (2023). ACCESS: Advancing innovation: NSF's advanced cyberinfrastructure coordination ecosystem: Services & support. In *Practice and experience in advanced research computing* (pp. 173–176). Association for Computing Machinery. https://doi.org/10.1145/3569951.3597559

Brown, S. T., Buitrago, P., Hanna, E., Sanielevici, S., Scibek, R., & Nystrom, N. A. (2021). Bridges-2: A platform for rapidly-evolving and data intensive research. In *Practice and experience in advanced research computing* (pp. 1–4). Association for Computing Machinery. https://doi.org/10.1145/3437359.3465593

Buckley, M. W., & Marshall, J. (2016). Observations, inferences, and mechanisms of the Atlantic meridional overturning circulation: A review. *Reviews of Geophysics*, *54*(1), 5–63. https://doi.org/10.1002/2015RG000493

Caesar, L., Rahmstorf, S., Robinson, A., Feulner, G., & Saba, V. (2018). Observed fingerprint of a weakening Atlantic Ocean overturning circulation. *Nature*, *556*(7700), 191–196. https://doi.org/10.1038/s41586-018-0006-5

Cessi, P. (2019). The global overturning circulation. *Annual Review of Marine Science*, *11*(1), 249–270. https://doi.org/10.1146/annurev-marine-010318-095241

Chang, C.-Y., & Jansen, M. F. (2022). The time-dependent response of a two-basin ocean to a sudden surface temperature change. *Journal of Climate*, *35*(14), 4535–4549. https://doi.org/10.1175/JCLI-D-21-0821.1

Cleveland, W. S. (1979). Robust locally weighted regression and smoothing scatterplots. *Journal of the American Statistical Association*, *74*(368), 829–836. https://doi.org/10.1080/01621459.1979.10481038

Cunningham, S. A., Kanzow, T., Rayner, D., Baringer, M. O., Johns, W. E., Marotzke, J., et al. (2007). Temporal variability of the Atlantic meridional overturning circulation at 26.5°N. *Science*, *317*(5840), 935–938. https://doi.org/10.1126/science.1141304

Duchez, A., Courtois, P., Harris, E., Josey, S. A., Kanzow, T., Marsh, R., et al. (2016). Potential for seasonal prediction of Atlantic sea surface temperatures using the RAPID array at 26°N. *Climate Dynamics*, *46*(9–10), 3351–3370. https://doi.org/10.1007/s00382-015-2918-1

Elipot, S., Frajka-Williams, E., Hughes, C. W., Olhede, S., & Lankhorst, M. (2017). Observed basin-scale response of the North Atlantic meridional overturning circulation to wind stress forcing. *Journal of Climate*, *30*(6), 2029–2054. https://doi.org/10.1175/JCLI-D-16-0664.1

Elipot, S., Frajka-Williams, E., Hughes, C. W., & Willis, J. K. (2014). The observed North Atlantic meridional overturning circulation: Its meridional coherence and ocean bottom pressure. *Journal of Physical Oceanography*, *44*(2), 517–537. https://doi.org/10.1175/JPO-D-13-026.1

Elipot, S., Hughes, C., Olhede, S., & Toole, J. (2013). Coherence of western boundary pressure at the RAPID WAVE array: Boundary wave adjustments or deep western boundary current advection? *Journal of Physical Oceanography*, *43*(4), 744–765. https://doi.org/10.1175/JPO-D-12-067.1

Ferrari, R., Griffies, S. M., Nurser, A. J. G., & Vallis, G. K. (2010). A boundary-value problem for the parameterized mesoscale eddy transport. *Ocean Modelling*, *32*(3–4), 143–156. https://doi.org/10.1016/j.ocemod.2010.01.004

Figa-Saldaña, J., Wilson, J. J. W., Attema, E., Gelsthorpe, R., Drinkwater, M. R., & Stoffelen, A. (2002). The advanced scatterometer (ASCAT) on the meteorological operational (MetOp) platform: A follow on for European wind scatterometers. *Canadian Journal of Remote Sensing*, *28*(3), 404–412. https://doi.org/10.5589/m02-035

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. https://doi.org/10.5194/gmd-8-3071-2015

Frajka-Williams, E. (2015). Estimating the Atlantic overturning at 26°N using satellite altimetry and cable measurements. *Geophysical Research Letters*, *42*(9), 3458–3464. https://doi.org/10.1002/2015GL063220

Frajka-Williams, E., Ansorge, I. J., Baehr, J., Bryden, H. L., Chidichimo, M. P., Cunningham, S. A., et al. (2019). Atlantic meridional overturning circulation: Observed transport and variability. *Frontiers in Marine Science*, *6*, 260. https://doi.org/10.3389/fmars.2019.00260

Frölicher, T. L., Sarmiento, J. L., Paynter, D. J., Dunne, J. P., Krasting, J. P., & Winton, M. (2015). Dominance of the Southern Ocean in anthropogenic carbon and heat uptake in CMIP5 models. *Journal of Climate*, *28*(2), 862–886. https://doi.org/10.1175/JCLI-D-14-00117.1

Fu, L.-L., Christensen, E. J., Yamarone Jr, C. A., Lefebvre, M., Menard, Y., Dorrer, M., & Escudier, P. (1994). TOPEX/POSEIDON mission overview. *Journal of Geophysical Research*, *99*(C12), 24369–24381. https://doi.org/10.1029/94JC01761

Fu, L.-L., Pavelsky, T., Crétaux, J.-F., Morrow, R., Farrar, J. T., Vaze, P., et al. (2024). The surface water and ocean topography mission: A breakthrough in radar remote sensing of the ocean and land surface water. *Geophysical Research Letters*, *51*(4), e2023GL107652. https://doi.org/10.1029/2023GL107652

Galbraith, E., & de Lavergne, C. (2019). Response of a comprehensive climate model to a broad range of external forcings: Relevance for deep ocean ventilation and the development of late Cenozoic ice ages. *Climate Dynamics*, *52*(1–2), 653–679. https://doi.org/10.1007/s00382-018-4157-8

Gent, P. R., & McWilliams, J. C. (1990). Isopycnal mixing in ocean circulation models. *Journal of Physical Oceanography*, *20*(1), 150–155. https://doi.org/10.1175/1520-0485(1990)020⟨0150:IMIOCM⟩2.0.CO;2

Gent, P. R., Willebrand, J., McDougall, T. J., & McWilliams, J. C. (1995). Parameterizing eddy-induced tracer transports in ocean circulation models. *Journal of Physical Oceanography*, *25*(4), 463–474. https://doi.org/10.1175/1520-0485(1995)025⟨0463:PEITTI⟩2.0.CO;2

George, T. M., Manucharyan, G. E., & Thompson, A. F. (2021). Deep learning to infer eddy heat fluxes from sea surface height patterns of mesoscale turbulence. *Nature Communications*, *12*(1), 800. https://doi.org/10.1038/s41467-020-20779-9

Griffies, S. M. (2012). Elements of the modular ocean model (MOM) [Computer software manual]. *GFDL Ocean Group Technical Report No. 7*. Retrieved from https://mom-ocean.github.io/assets/pdfs/MOM5_manual.pdf

Griffies, S. M., Danabasoglu, G., Durack, P. J., Adcroft, A. J., Balaji, V., Böning, C. W., et al. (2016). OMIP contribution to CMIP6: Experimental and diagnostic protocol for the physical component of the Ocean Model Intercomparison Project. *Geoscientific Model Development*, *9*(9), 3231–3296. https://doi.org/10.5194/gmd-9-3231-2016

Griffies, S. M., Gnanadesikan, A., Dixon, K. W., Dunne, J. P., Gerdes, R., Harrison, M. J., et al. (2005). Formulation of an ocean model for global climate simulations. *Ocean Science*, *1*(1), 45–79. https://doi.org/10.5194/os-1-45-2005

Gu, S., Liu, Z., & Wu, L. (2020). Time scale dependence of the meridional coherence of the Atlantic meridional overturning circulation. *Journal of Geophysical Research: Oceans*, *125*(3), e2019JC015838. https://doi.org/10.1029/2019JC015838

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR)* (pp. 770–778). IEEE. https://doi.org/10.1109/CVPR.2016.90

Held, I. M., Guo, H., Adcroft, A., Dunne, J. P., Horowitz, L. W., Krasting, J. P., et al. (2019). Structure and performance of GFDL's CM4.0 climate model. *Journal of Advances in Modeling Earth Systems*, *11*(11), 3691–3727. https://doi.org/10.1029/2019MS001829

Higginson, S., Thompson, K. R., Woodworth, P. L., & Hughes, C. W. (2015). The tilt of mean sea level along the east coast of North America. *Geophysical Research Letters*, *42*(5), 1471–1479. https://doi.org/10.1002/2015GL063186

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, *9*(8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, *12*(1), 55–67. https://doi.org/10.1080/00401706.1970.10488634

Hoffman, R. N., & Leidner, S. M. (2005). An introduction to the near–real–time QuikSCAT data. *Weather and Forecasting*, *20*(4), 476–493. https://doi.org/10.1175/WAF841.1

Hughes, C. W., Williams, J., Blaker, A., Coward, A., & Stepanov, V. (2018). A window on the deep ocean: The special value of ocean bottom pressure for monitoring the large-scale, deep-ocean circulation. *Progress in Oceanography*, *161*, 19–46. https://doi.org/10.1016/j.pocean.2018.01.011

Hunke, E. C., Lipscomb, W. H., Turner, A. K., Jeffery, N., & Elliott, S. (2010). CICE: The Los Alamos sea ice model documentation and software User's manual version 4.1 [Computer software manual]. *Los Alamos National Laboratory, LA-CC-06-012*. Retrieved from https://csdms.colorado.edu/w/images/CICE_documentation_and_software_user%27s_manual.pdf

Jackson, L. C., Hewitt, H. T., Bruciaferri, D., Calvert, D., Graham, T., Guiavarc'h, C., et al. (2023). Challenges simulating the AMOC in climate models. *Philosophical Transactions of the Royal Society A*, *381*(2262), 20220187. https://doi.org/10.1098/rsta.2022.0187

Jeffree, J., Hogg, A. M., Morrison, A. K., Solodoch, A., Stewart, A. L., & McGirr, R. (2024). Grace satellite observations of Antarctic bottom water transport variability. *Journal of Geophysical Research: Oceans*, *129*(10), e2024JC020990. https://doi.org/10.1029/2024JC020990

Johnson, H. L., & Marshall, D. P. (2002). A theory for the surface Atlantic response to thermohaline variability. *Journal of Physical Oceanography*, *32*(4), 1121–1132. https://doi.org/10.1175/1520-0485(2002)032⟨1121:ATFTSA⟩2.0.CO;2

Kerr, Y. H., Waldteufel, P., Wigneron, J.-P., Martinuzzi, J. M., Font, J., & Berger, M. (2001). Soil moisture retrieval from space: The soil moisture and ocean salinity (SMOS) mission. *IEEE Transactions on Geoscience and Remote Sensing*, *39*(8), 1729–1735. https://doi.org/10.1109/36.942551

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*. Retrieved from https://arxiv.org/abs/1412.6980

Kiss, A. E., Hogg, A. M., Hannah, N., Boeira Dias, F., Brassington, G. B., Chamberlain, M. A., et al. (2020). ACCESS-OM2 v1.0: A global ocean–sea ice model at three resolutions. *Geoscientific Model Development*, *13*(2), 401–442. https://doi.org/10.5194/gmd-13-401-2020

Kong, H., & Jansen, M. F. (2021). The impact of topography and eddy parameterization on the simulated Southern Ocean circulation response to changes in surface wind stress. *Journal of Physical Oceanography*, *51*(3), 825–843. https://doi.org/10.1175/JPO-D-20-0182.1

Kostov, Y., Messias, M.-J., Mercier, H., Johnson, H. L., & Marshall, D. P. (2023). Fast mechanisms linking the Labrador Sea with subtropical Atlantic overturning. *Climate Dynamics*, *60*(9), 2687–2712. https://doi.org/10.1007/s00382-022-06344-5

Krogh, A., & Hertz, J. (1991). A simple weight decay can improve generalization. In *Advances in neural information processing systems* (Vol. 4). Retrieved from https://proceedings.neurips.cc/paper/1991/file/8eefcfdf5990e441f0fb6f3fad709e21-Paper.pdf

Kusahara, K., Williams, G. D., Tamura, T., Massom, R., & Hasumi, H. (2017). Dense shelf water spreading from Antarctic coastal polynyas to the deep Southern Ocean: A regional circumpolar model study. *Journal of Geophysical Research: Oceans*, *122*(8), 6238–6253. https://doi.org/10.1002/2017JC012911

Lambin, J., Morrow, R., Fu, L.-L., Willis, J. K., Bonekamp, H., Lillibridge, J., et al. (2010). The OSTM/Jason-2 mission. *Marine Geodesy*, *33*(S1), 4–25. https://doi.org/10.1080/01490419.2010.491030

Landerer, F. W., Wiese, D. N., Bentel, K., Boening, C., & Watkins, M. M. (2015). North Atlantic meridional overturning circulation variations from GRACE ocean bottom pressure anomalies. *Geophysical Research Letters*, *42*(19), 8114–8121. https://doi.org/10.1002/2015GL065730

Little, C. M., Hu, A., Hughes, C. W., McCarthy, G. D., Piecuch, C. G., Ponte, R. M., & Thomas, M. D. (2019). The relationship between U.S. East Coast sea level and the Atlantic meridional overturning circulation: A review. *Journal of Geophysical Research: Oceans*, *124*(9), 6435–6458. https://doi.org/10.1029/2019JC015152

Little, C. M., Zhao, M., & Buckley, M. W. (2020). Do surface temperature indices reflect centennial-timescale trends in Atlantic meridional overturning circulation strength? *Geophysical Research Letters*, *47*(22), e2020GL090888. https://doi.org/10.1029/2020GL090888

Lozier, M. S., Bacon, S., Bower, A. S., Cunningham, S. A., de Jong, M. F., de Steur, L., et al. (2017). Overturning in the subpolar North Atlantic program: A new international ocean observing system. *Bulletin of the American Meteorological Society*, *98*(4), 737–752. https://doi.org/10.1175/BAMS-D-16-0057.1

Lozier, M. S., Li, F., Bacon, S., Bahr, F., Bower, A. S., Cunningham, S. A., et al. (2019). A sea change in our view of overturning in the subpolar North Atlantic. *Science*, *363*(6426), 516–521. https://doi.org/10.1126/science.aau6592

Lozier, M. S., Roussenov, V., Reed, M. S. C., & Williams, R. G. (2010). Opposing decadal changes for the North Atlantic meridional overturning circulation. *Nature Geoscience*, *3*(10), 728–734. https://doi.org/10.1038/ngeo947

Mackay, Q., Fan, Y., Karnauskas, K. B., & Li, L. (2024). Nonstationarity of the Atlantic meridional overturning circulation's fingerprint on sea surface temperature. *Geophysical Research Letters*, *51*(19), e2024GL109789. https://doi.org/10.1029/2024GL109789

Mak, J., Maddison, J. R., Marshall, D. P., & Munday, D. R. (2018). Implementation of a geometrically informed and energetically constrained mesoscale eddy parameterization in an ocean circulation model. *Journal of Physical Oceanography*, *48*(10), 2363–2382. https://doi.org/10.1175/JPO-D-17-0190.1

Manucharyan, G. E., Siegelman, L., & Klein, P. (2021). A deep learning approach to spatiotemporal sea surface height interpolation and estimation of deep currents in geostrophic ocean turbulence. *Journal of Advances in Modeling Earth Systems*, *13*(1), e2019MS001965. https://doi.org/10.1029/2019MS001965

Marshall, J., & Speer, K. (2012). Closure of the meridional overturning circulation through southern ocean upwelling. *Nature Geoscience*, *5*(3), 171–180. https://doi.org/10.1038/ngeo1391

Mazloff, M. R., & Boening, C. (2016). Rapid variability of Antarctic bottom water transport into the Pacific Ocean inferred from GRACE. *Geophysical Research Letters*, *43*(8), 3822–3829. https://doi.org/10.1002/2016GL068474

Mazloff, M. R., Heimbach, P., & Wunsch, C. (2010). An eddy-permitting southern ocean state estimate. *Journal of Physical Oceanography*, *40*(5), 880–899. https://doi.org/10.1175/2009JPO4236.1

McCarthy, G. D., Brown, P. J., Flagg, C. N., Goni, G., Houpert, L., Hughes, C. W., et al. (2020). Sustainable observations of the AMOC: Methodology and technology. *Reviews of Geophysics*, *58*(1), e2019RG000654. https://doi.org/10.1029/2019RG000654

McDonald, G. C. (2009). Ridge regression. *Wiley Interdisciplinary Reviews: Computational Statistics*, *1*(1), 93–100. https://doi.org/10.1002/wics.14

Meng, S., Stewart, A. L., & Manucharyan, G. (2024). Circumpolar transport and overturning strength inferred from satellite observables using deep learning in an eddying southern ocean channel model. *Journal of Advances in Modeling Earth Systems*, *16*(9), e2024MS004262. https://doi.org/10.1029/2024MS004262

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*(3–4), 306–313. https://doi.org/10.1016/j.cosust.2013.03.002

Michel, S., Dijkstra, H., Guardamagna, F., Jacques-Dumas, V., van Westen, R., & von der Heydt, A. (2023). Deep learning reconstruction of Atlantic Meridional Overturning Circulation strength validates ongoing twenty-first century decline. *PREPRINT (Version 1) available at Research Square*. https://doi.org/10.21203/rs.3.rs-2744951/v1

Morrison, A. K., & Hogg, A. M. (2013). On the relationship between southern ocean overturning and ACC transport. *Journal of Physical Oceanography*, *43*(1), 140–148. https://doi.org/10.1175/JPO-D-12-057.1

Morrison, A. K., Hogg, A. M., England, M. H., & Spence, P. (2020). Warm circumpolar deep water transport toward Antarctica driven by local dense water export in canyons. *Science Advances*, *6*(18), eaav2516. https://doi.org/10.1126/sciadv.aav2516

Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I—A discussion of principles. *Journal of Hydrology*, *10*(3), 282–290. https://doi.org/10.1016/0022-1694(70)90255-6

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., et al. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830. https://doi.org/10.5555/1953048.2078195

Pillar, H. R., Heimbach, P., Johnson, H. L., & Marshall, D. P. (2016). Dynamical attribution of recent variability in Atlantic overturning. *Journal of Climate*, *29*(9), 3339–3352. https://doi.org/10.1175/JCLI-D-15-0727.1

Purkey, S. G., Smethie Jr, W. M., Gebbie, G., Gordon, A. L., Sonnerup, R. E., Warner, M. J., & Bullister, J. L. (2018). A synoptic view of the ventilation and circulation of Antarctic bottom water from chlorofluorocarbons and natural tracers. *Annual Review of Marine Science*, *10*(1), 503–527. https://doi.org/10.1146/annurev-marine-121916-063414

Riahi, K., Van Vuuren, D. P., Kriegler, E., Edmonds, J., O'Neill, B. C., Fujimori, S., et al. (2017). The Shared Socioeconomic Pathways and their energy, land use, and greenhouse gas emissions implications: An overview. *Global Environmental Change*, *42*, 153–168. https://doi.org/10.1016/j.gloenvcha.2016.05.009

Rintoul, S. R. (2018). The global influence of localized dynamics in the Southern Ocean. *Nature*, *558*(7709), 209–218. https://doi.org/10.1038/s41586-018-0182-3

Roussenov, V. M., Williams, R. G., Hughes, C. W., & Bingham, R. J. (2008). Boundary wave communication of bottom pressure and overturning changes for the North Atlantic. *Journal of Geophysical Research*, *113*(C8), C08042. https://doi.org/10.1029/2007JC004501

Salomonson, V. V., Barnes, W. L., Maymon, P. W., Montgomery, H. E., & Ostrow, H. (1989). MODIS: Advanced facility instrument for studies of the Earth as a system. *IEEE Transactions on Geoscience and Remote Sensing*, *27*(2), 145–153. https://doi.org/10.1109/TGRS.1989.35996

Sanchez-Franks, A., Frajka-Williams, E., Moat, B. I., & Smeed, D. A. (2021). A dynamically based method for estimating the Atlantic meridional overturning circulation at 26°N from satellite altimetry. *Ocean Science*, *17*(5), 1321–1340. https://doi.org/10.5194/os-17-1321-2021

Send, U., Lankhorst, M., & Kanzow, T. (2011). Observation of decadal change in the Atlantic meridional overturning circulation using 10 years of continuous transport data. *Geophysical Research Letters*, *38*(24), L24606. https://doi.org/10.1029/2011GL049801

Solodoch, A., Stewart, A. L., Hogg, A. M., Morrison, A. K., Kiss, A. E., Thompson, A. F., et al. (2022). How does Antarctic bottom water cross the Southern Ocean? *Geophysical Research Letters*, *49*(7), e2021GL097211. https://doi.org/10.1029/2021GL097211

Solodoch, A., Stewart, A. L., M. 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. https://doi.org/10.1029/2022MS003370

Srinivasan, K., Chekroun, M. D., & McWilliams, J. C. (2024). Turbulence closure with small, local neural networks: Forced two-dimensional and β-plane flows. *Journal of Advances in Modeling Earth Systems*, *16*(4), e2023MS003795. https://doi.org/10.1029/2023MS003795

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, *15*(1), 1929–1958. https://doi.org/10.5555/2627435.2670313

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), e2021GL094569. https://doi.org/10.1029/2021GL094569

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. https://doi.org/10.1175/JPO-D-17-0007.1

Stone, M. (1974). Cross-validatory choice and assessment of statistical predictions. *Journal of the Royal Statistical Society: Series B*, *36*(2), 111–133. https://doi.org/10.1111/j.2517-6161.1974.tb00994.x

Sun, S., Eisenman, I., Zanna, L., & Stewart, A. L. (2020). Surface constraints on the depth of the Atlantic meridional overturning circulation: Southern ocean versus North Atlantic. *Journal of Climate*, *33*(8), 3125–3149. https://doi.org/10.1175/JCLI-D-19-0622.1

Sun, S., & Thompson, A. F. (2020). Centennial changes in the Indonesian Throughflow connected to the Atlantic meridional overturning circulation: The ocean's transient conveyor belt. *Geophysical Research Letters*, *47*(21), e2020GL090615. https://doi.org/10.1029/2020GL090615

Sun, S., Thompson, A. F., Xie, S.-P., & Long, S.-M. (2022). Indo-Pacific warming induced by a weakening of the Atlantic meridional overturning circulation. *Journal of Climate*, *35*(2), 815–832. https://doi.org/10.1175/JCLI-D-21-0346.1

Sun, S., Thompson, A. F., Yu, J., & Wu, L. (2024). Transient overturning changes cause an upper-ocean nutrient decline in a warming climate. *Nature Communications*, *15*(1), 7727. https://doi.org/10.1038/s41467-024-52200-0

Tapley, B. D., Bettadpur, S., Ries, J. C., Thompson, P. F., & Watkins, M. M. (2004). GRACE measurements of mass variability in the Earth system. *Science*, *305*(5683), 503–505. https://doi.org/10.1126/science.1099192

Toms, B. A., Barnes, E. A., & Ebert-Uphoff, I. (2020). Physically interpretable neural networks for the geosciences: Applications to Earth system variability. *Journal of Advances in Modeling Earth Systems*, *12*(9), e2019MS002002. https://doi.org/10.1029/2019MS002002

Vallis, G. K. (2017). *Atmospheric and oceanic fluid dynamics*. Cambridge University Press. https://doi.org/10.1017/9781107588417

van Sebille, E., Baringer, M. O., Johns, W. E., Meinen, C. S., Beal, L. M., de Jong, M. F., & van Aken, H. M. (2011). Propagation pathways of classical Labrador Sea water from its source region to 26°N. *Journal of Geophysical Research*, *116*(C12), C12027. https://doi.org/10.1029/2011JC007171

Wang, H., Legg, S. A., & Hallberg, R. W. (2015). Representations of the Nordic seas overflows and their large scale climate impact in coupled models. *Ocean Modelling*, *86*, 76–92. https://doi.org/10.1016/j.ocemod.2014.12.005

Weare, B. C., & Nasstrom, J. S. (1982). Examples of extended empirical orthogonal function analyses. *Monthly Weather Review*, *110*(6), 481–485. https://doi.org/10.1175/1520-0493(1982)110⟨0481:EOEEOF⟩2.0.CO;2

Wei, H. (2024). Python script for training the dual-branch neural network described in the manuscript "full-depth reconstruction of long-term meridional overturning circulation variability from satellite-measurable quantities via machine learning" (version 1) [Software]. *Zenodo*. https://doi.org/10.5281/zenodo.14233051

Wei, H. (2025). Matlab scripts and data used to generate the figures in the manuscript "full-depth reconstruction of long-term meridional overturning circulation variability from satellite-measurable quantities via machine learning". *Zenodo*. https://doi.org/10.5281/zenodo.15643649

Willis, J. K. (2010). Can in situ floats and satellite altimeters detect long-term changes in Atlantic Ocean overturning? *Geophysical Research Letters*, *37*(6), L06602. https://doi.org/10.1029/2010GL042372

Wunsch, C., & Heimbach, P. (2013). Two decades of the Atlantic meridional overturning circulation: Anatomy, variations, extremes, prediction, and overcoming its limitations. *Journal of Climate*, *26*(18), 7167–7186. https://doi.org/10.1175/JCLI-D-12-00478.1

Xie, C., Wei, H., & Wang, Y. (2023a). Bathymetry-aware mesoscale eddy parameterizations across upwelling slope fronts: A machine learning–augmented approach. *Journal of Physical Oceanography*, *53*(12), 2861–2891. https://doi.org/10.1175/JPO-D-23-0017.1

Xie, C., Wei, H., & Wang, Y. (2023b). Impact of parameterized isopycnal diffusivity on shelf-ocean exchanges under upwelling-favorable winds: Offline tracer simulations augmented by artificial neural network. *Journal of Advances in Modeling Earth Systems*, *15*(4), e2022MS003424. https://doi.org/10.1029/2022MS003424

Yeager, S. (2020). The abyssal origins of North Atlantic decadal predictability. *Climate Dynamics*, *55*(7), 2253–2271. https://doi.org/10.1007/s00382-020-05353-1

Yeager, S., Castruccio, F., Chang, P., Danabasoglu, G., Maroon, E., Small, J., et al. (2021). An outsized role for the Labrador Sea in the multidecadal variability of the Atlantic overturning circulation. *Science Advances*, *7*(41), eabh3592. https://doi.org/10.1126/sciadv.abh3592

Zhai, Z.-M., Moradi, M., Panahi, S., Wang, Z.-H., & Lai, Y.-C. (2024). Machine-learning nowcasting of the Atlantic meridional overturning circulation. *APL Machine Learning*, *2*(3). https://doi.org/10.1063/5.0207539

Zhang, R. (2008). Coherent surface-subsurface fingerprint of the Atlantic meridional overturning circulation. *Geophysical Research Letters*, *35*(20), L20705. https://doi.org/10.1029/2008GL035463

Zhang, R. (2010). Latitudinal dependence of Atlantic meridional overturning circulation (AMOC) variations. *Geophysical Research Letters*, *37*(16), L16703. https://doi.org/10.1029/2010GL044474

Zhang, R., Sutton, R., Danabasoglu, G., Kwon, Y.-O., Marsh, R., Yeager, S. G., et al. (2019). A review of the role of the Atlantic meridional overturning circulation in Atlantic multidecadal variability and associated climate impacts. *Reviews of Geophysics*, *57*(2), 316–375. https://doi.org/10.1029/2019RG000644

Ziehn, T., Chamberlain, M. A., Law, R. M., Lenton, A., Bodman, R. W., Dix, M., et al. (2020). The Australian Earth system model: Access-ESM1.5. *Journal of Southern Hemisphere Earth Systems Science*, *70*(1), 193–214. https://doi.org/10.1071/ES19035

Erratum

The originally published version of this article contained a typographical error. Figures 10 and 11 were swapped. The error has been corrected, and this may be considered the authoritative version of record.