Estimating the AMOC from Argo Profiles with Machine Learning Trained on Ocean Simulations

Yannick Wölker[1,2], Willi Rath[1], Matthias Renz[2], and Arne Biastoch[1,2]

[1]GEOMAR Helmholtz Centre for Ocean Research Kiel, Kiel, Germany
[2]Christian-Albrechts Universität zu Kiel, Kiel, Germany

Correspondence: Yannick Wölker (ywoelker@geomar.de)

Abstract. The Atlantic Meridional Overturning Circulation (AMOC) plays an important role in our climate system, continuous monitoring is important and could be enhanced by combining all available information. Moored measuring arrays like RAPID divide the AMOC in near-surface contributions, western-boundary currents, and the deep ocean in the interior of the basin. For the deep-ocean component, moorings measure density and focus on the calculation through geostrophy. These moored devices come with a high maintenance effort. Existing reconstruction studies show success with near-surface variables on monthly time scales, but do not focus on the interior transport. For interannual to decadal time scales, the geostrophic contribution becomes an important contribution.

Argo floats could provide required information about the geostrophic circulation as they continuously and cost-effective deliver hydrographic profiles. But they are spatially unstructured and only report instantaneous values. Here we show that the geostrophic part of the AMOC can be ~~data-drivenly~~ potentially reconstructed by Argo profiles in a data-driven methodology. To demonstrate this, we use a realistic and physically consistent high-resolution model VIKING20X. By simulating virtual Argo floats, we demonstrate that a learnable binning method to process the spatially variable Argo float distribution is able to reconstruct the geostrophic part of the VIKING20X AMOC by up to 80% explained variance and a mean error of less than one Sverdrup for the geostrophic transport. Using methods of explainable AI we investigate the importance of our input components showing an increasing importance of the Argo profiles on seasonal and interannual timescales, validating the usefulness of the Argo floats for the reconstruction. Our results demonstrate how an AMOC reconstruction from unstructured Argo profiles could replace estimates of the geostrophic deep-ocean component of the AMOC from the RAPID Array in the context of high-resolution ocean and climate models.

1 Introduction

The Atlantic Meridional Overturning Circulation (AMOC) is a key part of the climate system, and has specific socio-economic implications for Europe and North America (Zhang et al., 2019). It moves warm and salty water northward in the upper layers of the Atlantic Ocean, while colder and fresher waters flow southward at depth (Lozier, 2012). Owing to this vertical temperature contrast, the AMOC carries about 1.2 PW ($1PW = 10^{15}W$) in heat northward and regulates global and regional climate (Johns et al., 2023). The AMOC is usually described as the total overturning transport of water across a full section

of the Atlantic. Because of its importance, its strength and variability are closely watched, and the question of how it might change in the future is widely debated (Caesar et al., 2021; Latif et al., 2022; Terhaar et al., 2025; Baker et al., 2025). The AMOC has been studied using both observational data (McCarthy et al., 2012; Li et al., 2021) and numerical models (Biastoch et al., 2021; Petit et al., 2023). Both ocean and climate models often fail to accurately simulate the AMOC (McCarthy and Caesar, 2023), potential simulated AMOC trends are possibly due to artificial numerical choices such as those made to balance the freshwater budget (Behrens et al., 2013; Biastoch et al., 2021) or the inability of coupled climate models to simulate the correct decadal variability (Jackson et al., 2022).

As an integral quantity describing the combined effect of several current systems, the AMOC is constructed from several components representing different transport processes, including boundary currents, wind-driven Ekman transports at the surface, and slower-moving currents in the (deep) ocean interior (Moat et al.; Frajka-Williams et al., 2019). Further complication comes from the fact that each of these processes is governed by different timescales. While the AMOC strongly varies on daily to seasonal timescales, with even short-term reversals, interannual changes mask the variability, the decadal variability and potential long-term trends of the AMOC are most important for climate (Latif and Keenlyside, 2011).

A major effort to monitor the Atlantic Meridional Overturning Circulation (AMOC) is undertaken at 26.5°N through the international RAPID project (Cunningham et al., 2007; Johns et al., 2023). At this latitude, RAPID breaks the total overturning down into three main parts: ~~The first part is the Ekman transport, which is mainly driven by the zonal wind stress. It can be estimated from satellite-based wind stress products (McCarthy et al., 2015). The second part is the Florida Current, a strong western-boundary current, which is monitored using a~~ First, the wind-driven Ekman transport measured by satellites (McCarthy et al., 2015). Second, the strong Florida Current at the western boundary monitored by submarine telephone cables ~~that measurement changes through induction and electrical resistance and provides~~ between Florida and the Bahamas. Measurements of changes in electrical resistance provide a continuous time series at a daily resolution (Larsen, 2001). These two components are routinely available and can be directly included to construct the ~~total~~ AMOC. The third part, the upper mid-ocean transport, is more difficult to obtain. ~~Within RAPID, the~~ The mid-ocean transport is separated further into ~~different components. One is the Antilles Current (also known as the western-boundary wedge), which is measured using moored current meters east of the Antilles section (Johns et al., 2008). Another is the~~ the western boundary wedge (Johns et al., 2008), and the interior geostrophic transport, estimated from moorings ~~placed at both~~ at the eastern and western boundaries ~~of the section (over some periods also on the eastern and western sides of the Mid-Atlantic Ridge) and calculated mainly through thermal wind balance(McCarthy et al., 2015).~~ (McCarthy et al., 2015). In addition, a baroclinic transport term is included as a closure to ensure net-zero mass transport across the section. Establishing and maintaining such a mooring system is costly and depends on major international collaboration. Since measurements of the ocean interior are limited in time and location, the question arises of whether the geostrophic transport in the ocean interior can be estimated with less effort and data that is more widely available, and also for other latitudes.

This study explores the ~~usage of~~ concept of using passively drifting autonomous observing systems, such as ~~Argo floats~~ simulated Argo floats from ocean models, as an alternative for measuring the water mass structure and the derivation of geostrophic gradients in the upper 2000 meters ~~(?Roemmich et al., 2009)~~ (Wong et al., 2020; Roemmich et al., 2009). Argo floats offer a

60 cost-effective way to observe the ocean interior~~without being tied to a fixed latitude. They are promising measurements of~~ . They have potential for the upper part ~~from~~ of the RAPID moorings at different time scales, from a single ~~dive cycle of about 10 days~~10-day dive cycle, to mesoscale ~~patterns~~ times scales over 30 days, seasonal signals over 90 days, and even longer time scales. ~~Each~~ The Argo float cycle involves drifting at a depth of around 1,000 meters ~~. Every approximately 10 days, it dives to a depth of 2,000 meters~~ and a 2000 meter dive before returning to the surface~~, during which the float records a vertical~~

65 ~~profile of temperature and salinity. Data are transmitted through a satellite connection while the float drifts at the surface for a few hours. This system provides continuous hydrographic data across the upper ocean, covering wide parts of all Ocean basins. As a result,~~ . Argo floats have strong potential for estimating basin-wide transport because they provide continuous hydrographic data across the upper ocean. However, their irregular timing and spatial coverage make it challenging to apply standard reconstruction methods without information loss.

70 Earlier attempts to estimate the AMOC often relied on satellite-based measurements of surface quantities such as sea surface height (SSH), wind stress, or on gravimetric estimates of the Ocean bottom pressure (OBP). How well these methods perform generally depends on the timescale considered. For example, Frajka-Williams (2015) estimated the AMOC at 26°N using SSH together with estimates of the Florida Current and Ekman transport, capturing up to 90% of the interannual variability. However, unlike the RAPID approach, Frajka-Williams and others like Sanchez-Franks et al. (2021) do not use direct measurements but

75 rather quantities which are observable at the surface to approximate the ocean interior. Other studies have included subsurface data to improve the reconstruction of geostrophic transport. For instance, Solodoch et al. (2023) showed that bottom pressure and zonal wind stress in the ocean reanalysis ECCOv4 (Forget et al., 2015) are suitable to reconstruct the AMOC at this latitude using a simple neural network. However, since bottom pressure measurements are not widely available, Argo floats appear to be an attractive alternative, even though they do not deliver direct transport measurements. ~~Zilberman et al. (2020) grouped~~

80 ~~Argo profiles into 6°×6° cells in the Pacific to create a uniform coverage of Argo profiles which could be used for further computation. Also, Argo floats were used to map the density measurements onto a transport section (e.g., in the Agulhas region (McMonigal et al., 2018), and the Labrador Sea (Holte and Straneo, 2017)) to create seasonal means of the transport in the upper 2,000 meters based on the geostrophic balance. In Willis (2010) the authors computed~~ Multiple studies were conducted to average Argo profiles in regular cells (Zilberman et al., 2020; Willis, 2010), from which dynamic height fields ~~from Argo~~

85 ~~data~~ can be used to estimate interior geostrophic transports~~around 41°N. Build on those estimated transport Elipot et al. (2014) compared these to an estimate derived from the OBP gradient~~ (McMonigal et al., 2018; Holte and Straneo, 2017). A comparison between the estimated Argo transports and the OBP gradients at the western continental shelf ~~, finding that the major driver of the overturning at 26°N and 41°N is the western geostrophic transport. In a later work those transport series at 26°N and 41°N were used to estimate the analytic correlation to the wind component being able to reconstruct up to 50% variance of~~

90 ~~the transport series (Elipot et al., 2017).~~ was performed by Elipot et al. (2014). So far, these approaches rely on spatial and temporal binning to organize the Argo data, and average over multiple profiles to produce grid-based fields for further computation. Typically, three-month windows and spatial binning (e.g., quarter-degree) is used, which can smooth out important details and reduce the value of individual profiles. ~~There remains a gap in combining modern machine-learning approaches (Solodoch et al., 2023) with Argo float data. Most machine-learning approaches require structured input data with consistent~~

95 ~~spatial coverage and distribution, but Argo floats drift with the ocean currents and produce measurements that are irregular both in space and time.~~ None of the mentioned approaches fully preserves the detail in individual profiles, especially on short timescales where data sparsity becomes a major challenge. There remains a gap in combining modern machine-learning approaches (Solodoch et al., 2023) with Argo float data.

Machine learning methods for reconstructing the AMOC rely on supervised learning, ~~which means they need a lot of~~
100 requiring paired input-output data for training. Even though the RAPID array has been collecting data for almost 20 years (Johns et al., 2023), these data are not enough to train a reliable reconstruction, especially when targeting timescales longer than a month. For example, splitting the 20 years of RAPID data into 90-day averages would only yield around 80 independent samples~~, which is too few for a more complex machine learning method like a graph-based neural network to properly learn the spatial relationships between Argo profiles.~~.Hence, observational data alone cannot be used. ~~Furthermore, since~~ Since
105 observational datasets are limited in their coverage in time, we ~~could~~ can not rely on them to assess the generalizability ~~of a machine-learning approaches~~ to periods beyond the training data. To work around the issue of data scarcity, ~~some~~ earlier studies (Solodoch et al., 2023; Sonnewald and Lguensat, 2021) used ocean reanalysis products like ECCOv4~~, which offer longer time series on monthly scales.~~.Ocean reanalyses and ocean general circulation models provide longer and fully spatial and temporal covered datasets for training. While these ocean models do have their own biases and uncertainties, they are generally
110 consistent in terms of the underlying physical principles formulated in the hydrodynamic equations and how different parts of the AMOC respond to surface forcing (Biastoch et al., 2021; Böning et al., 2023). In particular, high-resolution ocean models have been proven to simulate the important aspects of the wind-driven and thermohaline circulation, including western boundary current and mesoscale dynamics, deep formation and spreading of water masses (Hirschi et al., 2020) This makes them a solid basis for learning heterogeneous ocean states with ~~known~~ inputs like wind stress, Florida Current, Antilles Current, and
115 hydrographic conditions ~~to reconstruct the AMOC strength as the output~~ (see Figure 1). Despite remaining errors due to numerical choices and parameterizations, models can be used as a well-defined testbed to develop and evaluate machine learning approaches for AMOC reconstruction. They offer a controlled but realistic setting ~~where we can test how well Argo-based reconstructions capture the full AMOC signal, including interior variability.~~ in which the data quality can be similar to the real world observations but with a increased temporal extent.

120 In this work, we demonstrate how and to what extent the AMOC can be reconstructed ~~from~~ in an ocean model setting from simulated measurements that mimic widely available observational products~~using supervised machine learning.~~.Reconstructing the AMOC on different time scales from 10 days up to five years, we find that the importance of geostrophic transport in the ocean interior becomes more pronounced, with longer time scales. To capture this, we use virtual Argo profiles as our observational input, leveraging their insight into the ocean interior while addressing their sparse and irregular sampling with a
125 graph-based neural network approach. ~~As the Argo profiles only cover the upper 2000m of the ocean, we also use positions of the RAPID moorings for information about the deeper layers. In addition, we evaluate the potential benefit of using Deep Argo floats reaching up to 6000m depth, as it is discussed to establish a permanent fleet of 1228 Deep Argo floats hence achieving an average distribution of one float per 5° x 5° spatial bin(Johnson et al., 2015).~~ To overcome the irregular spacing, we adapt SUSTeR (Wölker et al., 2023), an approach for reconstructing spatio-temporal events from unstructured sparse observations

spread over a wide area using graph neural networks. We extend this approach, originally developed for traffic reconstruction, to handle drifting sensors like Argo floats ~~by learning general concepts across space and time.Each Argo profile is treated as an individual input linked to one of those concepts, allowing the framework to dynamically group Argo profiles without relying on a static clustering.~~ .We train and evaluate our method with virtual Argo floats and virtual deep RAPID moorings sampled from the high-resolution ocean simulation VIKING20X (Biastoch et al., 2021).

While our main goal is to train a reconstruction for the ~~total~~ AMOC at 26.5°N based on direct measurements of Florida Current transport, Antilles Current transport, on wind stress estimates, and on Argo float and mooring data, we also investigate other experiments validating our framework by highlighting different components of the AMOC. To this end we not only take the ~~total~~ AMOC from VIKING20X as training target, but additionally generate a RAPID-like ~~geostrophic AMOC~~ interior geostrophic transport time series from virtual moorings. ~~All inputs are extracted from the ocean simulation together with virtual Argo floats which have the same spatio-temporal distribution as in the real-world, and then tested against the total AMOC calculated on the same simulation.~~ Our method successfully reconstructs the AMOC at 26.5°N on seasonal and longer timescales, largely due to its ability to recover interior geostrophic transport from Argo float profiles.
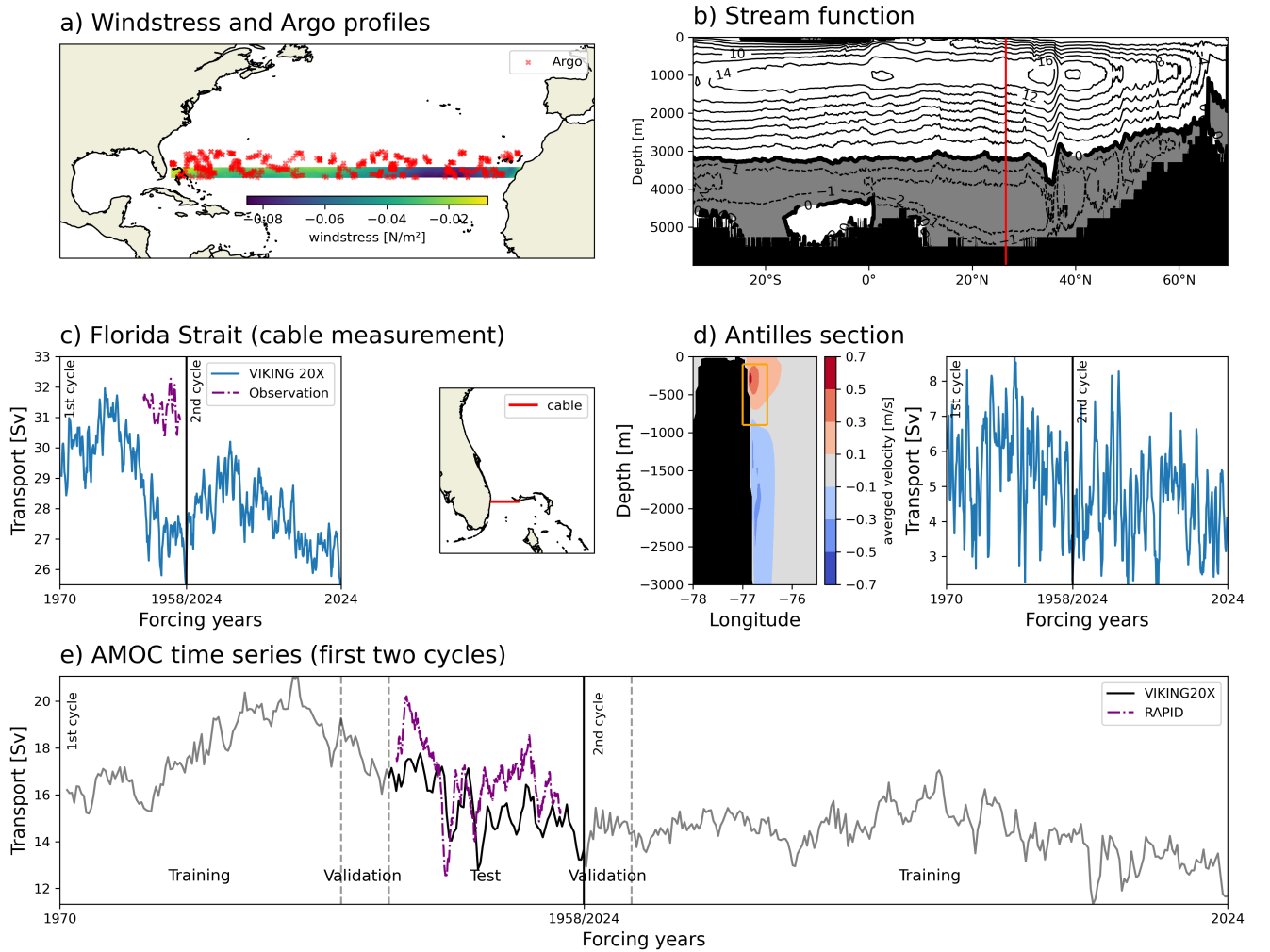
Figure 1. Sketch of all modalities that are included in the reconstruction of the AMOC in the context of the VIKING20X simulation. a) shows where wind stress around the target latitude was used for a sample season in 2015 together with all Argo position that were available for this 90 days period. The overturning stream function in the Atlantic basin b) in the reconstruction period from 2005 to 2022 shows the target latitude with a red line and the average structure of the overturning. The panel c) and d) show the Florida Current and the Antilles Current respectively. The left sub-panel shows the location of the measurement and the right sub-panel shows the resulting seasonal-mean transport time series used as input for the reconstruction. e) shows the ~~total~~ AMOC time series along the 26.5°N latitude from the first two cycles in the VIKING20X simulation which we use as training target for the reconstruction.

2 Data

2.1 Simulation Cycles of the eddy-rich ocean model VIKING20X

145 VIKING20X is an ocean general circulation model configuration with a high-resolution nest in the Atlantic basin. The nest provides eddy-rich resolution not only for the region of interest around the latitude 26.5°N but also at all relevant key regions for AMOC change. The model configuration is based on the NEMO framework and has a global OCRA025 configuration (1/4° resolution) with a higher resolution nest (1/20° resolution) covering the Atlantic from 33.5°S to about 65°N using Adaptive Grid Refinement (AGRIF) (Debreu et al., 2008). In both, the base and the refined nest component, the configuration has 46

150 vertical z-levels with a vertical resolution decreasing from 6m close to the surface to 250m close to the ocean floor. The model is subject to the JRA55-do atmospheric forcing between 1958 and 2023 (Tsujino et al., 2018). In alignment with the OMIP-II protocol (Tsujino et al., 2020), six cycles of the model simulation each covering the period of 1958 to 2023 were sequentially executed with the same forcing for each cycle. The first cycle was initialized from an ocean at rest and from hydrographic conditions reflecting the World Ocean Atlas 2013 (WOA13, Locarnini et al. (2013), Zweng et al. (2013)). Subsequent cycles

155 were started from the last time step of the previous cycle with a restarted forcing period. To account for a spinup time of the model, we only used the first cycle from 1970 onward, ~~the task we are training does not require a completely spun up ocean state because the variability in the training data can improve generalization as long as the gap to the other training data is not too large.~~. The high-resolution simulation VIKING20X is known to resolve mesoscale processes in the entire North Atlantic which includes key processes for the variability of the AMOC such as deep water formation and the process of deep

160 convection observable in the subpolar North Atlantic (Rühs et al., 2021), the distribution of fresh water in the subpolar gyre (Fox et al., 2022), and the overflow pattern (Böning et al., 2023), as well as aspects of the wind-driven circulation such as western boundary currents and mesoscale eddies (Biastoch et al., 2021).

  The AMOC strength and variability in the first cycle is described in detail in (Biastoch et al., 2021) and is used in this study as the reconstruction target. To increase the amount of data that can be utilized for the training ~~of the reconstruction~~process,

165 we also use virtual measurements from the second cycle. The AMOC strength for the first two cycles is shown in Figure ~~??~~ 1 e) on a seasonal time scale with 15.5Sv mean transport and 2.6Sv standard deviation. While the first of the VIKING20X cycles has a stronger decadal AMOC variability associated with the deepwater formation in the 1990s in the subpolar North Atlantic (Böning et al., 2023), the second cycle simulates a reduced decadal AMOC variability (Schiller-Weiss et al., 2025). We expect that the generalizability of our trained reconstruction would profit from an increased heterogeneity of the training

170 data as long as the training data and training targets represent consistent physical relations. Hence, even though the weakening decadal variability of the AMOC over the course of the six OMIP cycles may be due to spurious model drift, it may still be an advantage to include data from all six OMIP cycles into our training.

2.2 RAPID array

The RAPID Array is a multi-national effort to monitor the overturning at 26.5°N and is deployed since April 2004. ~~(~~(Kanzow

175 et al., 2008; McCarthy et al., 2015; Frajka-Williams et al., 2019)~~)~~RAPID estimates the AMOC in three different contributions,

the Ekman transport with remote sensing product~~((McCarthy et al., 2015))~~ (McCarthy et al., 2015), the Florida Current transport via an underwater cable in the Florida Strait (Larsen, 2001), and the upper mid-ocean component. This upper mid-ocean can be divided into the Antilles Current measured by moored current meters (Johns et al., 2008) and the interior geostrophic transport which is measured by moorings over large distances between the continental margins ~~and over some periods also at the mid oceanic ridge of the Atlantic basin (Kanzow et al., 2008; Sinha et al., 2018). For our study we mimic the RAPID moorings, that cover the basin interior transport, in the VIKING20X simulation and calculate the vertical gradient of the meridional velocity as the values of the so called virtual mooring. As the RAPID moorings do not cover the entire water column up to the surface we remove the surface-near values with a mask that is sampled from~~ (Kanzow et al., 2008; Sinha et al., 2018). We will use the interior geostrophic transport calculation from RAPID in the ~~real-world RAPID moorings and extrapolate the remaining values to the surface as it was done in the RAPID calculation (McCarthy et al., 2015). Those virtual moorings are used to calculate the geostrophic transport which we utilize in section 4.3 to investigate the ability to reconstruct the geostrophic transport only. This allows to have a similar information coverage as RAPID to investigate how much AMOC~~ context of the ocean simulation to investigate how much geostrophic variability can be reconstructed on different time scales with ~~only widely available information and observation-like components~~Argo floats.

2.3 Argo floats

Argo is an international network of Lagrangian floats which reside and drift with the ocean currents at 1000m depth and profile the upper 2000m of the water column in a fixed interval of 10 days (Riser et al., 2016). Argo floats provide the methodology to measure the interior of the ocean without any fixed mooring or recurrent ship survey~~but .~~ However, coming with the cost of a sparse and unstructured spatial distribution. Argo floats were first deployed in ~~the early 2000s~~1999 and their coverage has continuously grown until 2012 ~~when they reached their current operational state with an average of~~with approximately 430 Argo profiles per season in our study area.

Previous works averaged Argo measurements along larger time frames, Willis (2010); Hernández-Guerra et al. (2010) creating zonal-depth section of temperature and salinity profiles from which the geostrophic transport is inferred. Also the velocity information of the Argo floats at their parking depth, estimated by the offset of the surface position, was used to estimate ~~the~~ transport (Reeve et al., 2019; Zilberman et al., 2020, 2023)~~, which is not part of our reconstruction because the fixed parking depth limits the applicability of the presented approach to this depth.~~. Those methods neglect any ~~finer~~ individual float correlations due to the limitation of unstructured profiles~~and have to use interpolation methods for regions without Argos, which can lead to accumulated errors.~~. Solodoch et al. (2023), who reconstructed the AMOC with bottom pressure used for information about the ocean interior, highlighted the need to investigate how Argo floats can be adopted for this task. To handle the spatial distribution of ~~Argos~~Argo profiles, we here use a graph data structure (Dale and Fortin, 2010), which allows for a flexible topology~~, and each graph node representing a learned variable grained cluster of Argo profiles, as we explain further in section 3.1.~~.

2.4 Input & Output data for AMOC Reconstruction

While this study investigates to which extent observation strategies from the RAPID mooring array can be explained by Argo float data, it cannot directly use the observations from the real ocean, because the RAPID array only covers a period of 20 years. The resulting amount of independent samples from 20 years of observational data is not sufficient for training our machine learning framework, on longer than monthly or seasonal timescales. A supervised machine-learning approach is a general function approximation which requires data samples containing an input, the widely available measurements, and a target variable, the interior geostrophic or AMOC transport, to learn the reconstruction. In the following sections and summarized in Fig. 2 we describe in detail how the input and output is extracted from the ocean simulation.

Observation-like input products

The AMOC reconstruction is based on zonal wind stress, Florida Current transport, Antilles Current transport, Argo profiles for the upper 2000 meters, and the RAPID mooring information below (all components shown in Figure 1). To estimate the wind stress which in the RAPID approach is based on a remote sensing product, we take the zonal wind stress field from the forcing (calculated through Bulk formulae from the velocity difference between wind and surface ocean velocity) between 25°N and 27°N (Figure 1 a)) and create 10 zonal bins. Although the Ekman transport could be estimated based on zonally averaged zonal wind stress alone, we provide 10 bins of averaged zonal wind stress to retain some information about the spatial structure. The transport through the Florida Strait was calculated as the integrated volume transport that is discussed in further detail in Biastoch et al. (2021) for the first OMIP cycle of the ocean model VIKING20X. It is consistent with the observations (about 2Sv or 7% weaker) and interannually correlated (~ 0.75) with the AMOC. Inputs for the mid-ocean RAPID component are split into the Antilles Current and the interior of the ocean. For the Antilles Current we utilize a small box based on the mean transport pattern in VIKING20X in figure 1 d) from which we extract the Antilles Current transport. For the first two cycles in the Figure 1 d) the Antilles Current has an average transport of 5.5 Sv and 2.5 Sv standard deviation for seasonal time scales. This is meant to summarize the dynamics in the ocean state for everything that is west to the WB3 mooring in the RAPID array marking the boundary between the interior transport and the western boundary current. (Johns et al., 2011) For the region east of mooring WB3, we investigate the suitability of the Argo profiles as an alternative for the RAPID moorings in the upper 2000 meters. The density of the virtual Argo profiles are used in the Argo processing module of the reconstruction with the goal to identify the observed ocean state. Additionally, we provide the stream function below 2000 meters
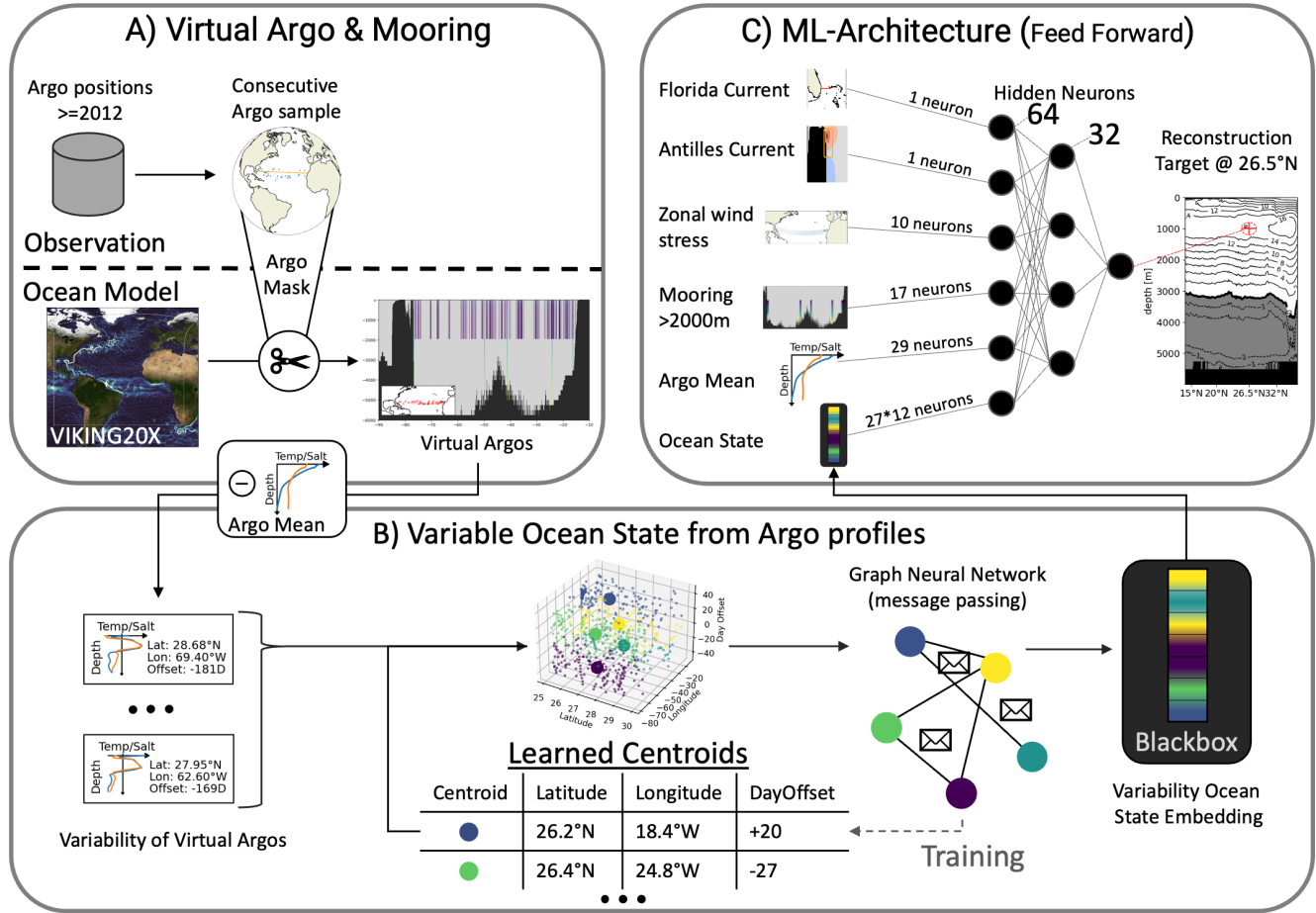
Figure 2. The workflow of the AMOC reconstruction with Argo profiles. A) shows the process of creating virtual Argo profiles from observational positions and relative times but with data extracted from an ocean simulation. The Argo profiles are divided for each time window in the mean profile and the residuals which are both used separately from each other. The variability is used in part B) of the figure to estimate the variable ocean state which is an embedding vector, also interpreted as black box, due to the end-to-end training. The reconstruction utilizes the strong power of GNNs to exchange information between learned cluster centers that aggregate single Argo profiles. The Argo mean, the estimated ocean states, and the inputs from Figure 1 are used in a feed-forward neural network to reconstruct the strength of the AMOC at 26.5°N and 1000 meter depth.

~~as information deeper than 2000 meters.~~ calculated from virtual RAPID moorings as an input to the reconstruction. Please note that the inputs to our neural network must not have the same physical units, since the learning process of the neural networks is based on the statistics of the input variables and not on their semantics. In section 4.4, we investigate the effectiveness of deep information by first testing the removal of the additional mooring data. Next, we assume that all Argo floats would reach up to 6000 meters.

We call the Argo floats and moorings, that we extract from VIKING20X, *virtual* to make clear that we mimic the spatial and temporal distribution but take the values from the simulation output (Figure 2 A). Therefore, we extract the ~~vertical grid boxes~~ horizontal grid box of the VIKING20X simulation nearest to the real Argo location from the daily output. To ensure the virtual profiles have the same spatio-temporal characteristics (and deficits) as the real Argo profiles, their position and time is sampled from the real-world Argo distribution. ~~A key characteristic of the real-world profiles~~ One example of such deficits is that consecutive real-world profiles from a single float are not independent from each other and that hence a uniform sampling across the Atlantic would grant too many degrees of freedom. To mitigate this we take random but consecutive time windows from the real world Argo distribution~~with equal length to the targeted reconstruction time scale. Each of these consecutive time windows will have virtual Argo profiles at the exact same positions and relative times as they were present in the real-world~~. The corresponding real-world time window of Argo profiles is sampled with uniform probability from all time windows of the respective length in the Atlantic between 25°N and 30°N~~and between~~, from 2012 ~~and 2022.~~

~~We do not include any real-world Argo distributions before 2012, because~~ (when the Argo array ~~which was only initiated in the early 2000s (Wong et al., 2020) showed an increasing deployment behavior starting at 2012. In the recent decade a constant number of Argo floats report values in our study region between 25°N and 30°N, which we assume to be the case in the future.~~ was full established) until the end of our simulation period 2023.

The processed virtual Argo product then provides these variables at vertical intervals of ~~20dbar~~20 dbar, as it is the standard for the real-world Argo floats. The ocean model from which we sample the virtual profiles ~~, on the other hand,~~ has a higher resolution at the surface (starting with 6m-thick grid cells) and coarser resolution ~~towards the deep ends of the profiles~~ at depth (208m at 2000m depth). ~~As we want to mimic the characteristics and deficits of the real-world profiles with our virtual profiles as closely as possible, we~~ We linearly interpolate the 46 z-levels of VIKING20X onto the depth resolution of the real-world Argo profiles between 20dbar to 2000dbar. Since Deep Argo (Roemmich et al., 2019) increases the number of floats ranging to depths close to the ocean floor, we also include a section 4.4 where we sample virtual ~~Argo profiles down to depths of 6000m to simulate a world with sufficient coverage by full-depth coverage by Deep Argo floats.~~ Deep Argo profiles.

AMOC strength target time series (Model & RAPID-like)

In this study, we investigate the reconstruction performance with two different ground-truth AMOC time series which are both calculated as the strength of the ~~AMOC at the nearest to~~ stream function at the grid box closest to 1000m ~~depth grid box~~ in the ocean model. First, we use the zonally integrated AMOC transport~~that is~~, calculated from a basin-wide cross section of the meridional velocities~~for all grid boxes in the nested configuration of VIKING20X. We call this AMOC timeseries the *total AMOC in the context of this work.*~~. This time series will cover all three main RAPID components (Ekman, upper mid-ocean,
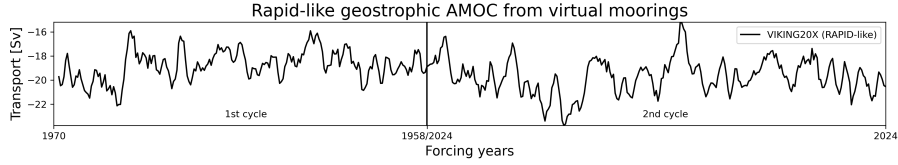
Figure 3. Showing the complete time series of the RAPID-like ~~geostorphic computed AMOC time series~~ interior geostrophic transport from virtual moorings extracted from the ocean model VIKING20X. This data is the target variable for the *geostrophic* ~~AMOC~~ experiments and shown here with a season low pass filter(90 days).

275 and Florida Strait)~~that were outlined by the RAPID array~~ (Frajka-Williams et al., 2019). Second, we put a strong focus on the predictive skill of Argo profiles for the ~~geostrophic transport~~ interior geostrophic transport (Figure 3) and their potential to be an alternative strategy for moored instruments~~, we also use a *geostrophic* AMOC time series in this work (Figure 3)~~. For the interior geostrophic transport, we calculate the RAPID upper mid-ocean transport and exclude the western boundary wedge, therefore this is similar to the gyre recirculation in Smeed et al. (2018). To be as similar as possible to the RAPID

280 ~~upper mid-ocean time series~~ interior transport, we took the RAPID calculation for the RAPID moorings (McCarthy et al., 2015) and mapped this to the VIKING20X output to share as many assumptions as possible. ~~Sinha et al. (2018) has shown that the RAPID methodology underestimates the mean AMOC in an eddy resolving ocean model slightly but represents the variability to an high accuracy, which motivates the usage of such an dedicated RAPID-like transport creating a baseline to the real-world RAPID time series. Therefore we use virtual moorings with a similar data loss and depth resolution to the real~~

285 ~~observations and calculated the northward geostrophic transport. We extract the strength of this aggregated stream function at 1000 meter depth to avoid the rule of the RAPID array, selecting the depth dependent on the presence of Antarctic Intermediate Water (AAIW) between 700m and 1000m (McCarthy et al., 2015). However, in a direct comparison to the Rapid timeseries two reconstructions can be trained one for 1000 meter and another for 700 meter. We argue that such an extraction following RAPID as close as possible is necessary to create an accurate time series that contains all the characteristics and caveats of~~

290 ~~the observations and to train the AMOC reconstruction most realistically to find reasonable correlations. A similar computed AMOC series is a strong prerequisite for any assumptions about transferability between the ocean model VIKING20X and the real-world observations within RAPID.~~ For a detailed description we refer the reader to appendix A.

Time scales

The AMOC is an integrated variable that compiles a number of physical processes on different time scales. For further insights,
295 we perform the AMOC reconstruction on different time scales from very short time scales of 10-days, with expected dominance of the geostrophic turbulence, through yearly timescales dominated by the wind-driven circulation, and beyond to observe the shift from the wind-driven variability to the ~~buyoancy-driven~~ buoyancy-driven ocean transports (Kanzow et al., 2010). The used range of timescales for each of the experiments, as well as the input parameters and the target variable for the training and testing period is documented for overview in Table 1. For a given time scale, the AMOC strength, the Florida Current,

Experiment	Input	Output	Scales	Research question
Reconstruction (Exp.1)	FC, AC, ZW, AR + MO	~~Total~~ AMOC	10d-1y	To which extent can the ground truth AMOC of VIKING20X be reconstructed?
Extended reconstruction (Exp.2)	FC, AC, ZW, AR + MO	~~Total~~ AMOC	90d-5y	Is the method able to generalize to an unseen cycle?
Component importance (Exp.3)	All but one: FC, AC, ZW, AR + MO	~~Total~~ AMOC	10d-1y	How does the impact of each input changes with the time scales
Geostrophy Isolation (Exp.4)	FC, AC, ZW, AR + MO	RAPID-like (geos)	10d-1y	Is the method able to reconstruct the geostrophic signal?
Deep Argo (Exp.5)	FC, AC, ~~WS~~ZW, and one of: AR, DAR, or AR+MO	~~Total~~ AMOC	10d to 1y	Which kind of information is needed below 2,000 m?

Table 1. For each experiment this table lists the used input features (FC - Florida Current, AC - Antilles Current, ZW - Zonal wind stress, AR+MO virtual Argo profiles <2,000m depth and virtual moorings >2,000m depth, AR - only virtual Argo profiles, DAR - virtual Deep Argo profiles) and which ~~AMOC~~ transport signal was set as the target (~~total~~ either the AMOC from Figure 1 e) or our self-computed ~~RAPID like geostrophic~~ RAPID-like interior geostrophic transport in panel f)). The time scales on which the experiment was evaluated. Additional the short research questions summarize the line of arguments in this work.

300 Antilles Current, and the binned zonal wind stress, are averaged to the target time interval based on daily ocean model output (e.g. the monthly averaged time series in Figure 1). The virtual Argo profiles ~~which represent instantaneous measurements~~ representing instantaneous measurements and can not be ~~averaged and are therefore concatenated~~ temporally averaged. Hence, we concatenate them into larger sets ~~of Argos~~ when the time scale is increasing. ~~This leads up to 90 day reconstructions with a maximum of up to 690 Argo profiles and 317 profiles on average. We therefore expect that the training on longer timescales~~

305 ~~requires a larger training set due to the larger number of Argo profiles in a single data point while the amount of available training data points decreases.~~

3 Method

The ~~reconstruction~~ machine-learning based econstruction of the AMOC ~~as a machine learning task~~ averaged to the respective target time scale, can be summarized as learning a set of parameters~~such that the trained reconstruction is able to reconstruct~~

310 ~~from an average~~, represented by a trained reconstruction model (in the following referred to as trained reconstruction), which uses a time-averaged zonal wind stress, ~~an averaged~~ a time-averaged Florida Current, ~~an averaged~~ a time-averaged Antilles Current, and a set of Argo profiles , for the same averaging period, to reconstruct the strength of the ~~averaged total or geostrophic AMOC~~ time-averaged AMOC or interior geostrophic transport at the depth of 1000 meters. The overall workflow of the method

is shown in Figure 2 with A) ~~being~~ depicting the creation of virtual ~~Argos~~Argo profiles as described in section 2.4, B) showing the processing of ~~those Argos to embed them in a hidden space~~this variable set of Argo profiles into a fixed-sized vector representation, and C) utilizing the ~~Argo embedding of the previous step together with the mean Argo profile from the first step~~ fixed-sized vector representation from step B) together with directly used estimates of the other contributions to estimate the AMOC strength. The architecture of the neural network has a dedicated Argo processing module (Figure 2 B), because the mapping from a variable amount of ~~Argos in each data sample and their sparseness require more advanced neural networks to find an suitable data structure~~spatially unstructured Argo profiles into a fixed-sized vector (black box in Figure 2 B) requires a more advanced method thatn for the other already structured contributions. We utilize artificial neural networks as our machine-learning framework to learn the reconstruction task due to their ability to approximate non-linear correlation in the unstructured spatio-temporal Argo profile distribution.

3.1 Processing of Argo profiles

~~For the virtual Argo profiles , the goal is to train an embedding (black box in Figure 2 B)) that maps a set of Argo profiles into a hidden space in where similar ocean states are near each other even though their spatial distribution of observations may be different. Note that an embedding here is a high-dimensional mathematical representation of the processed Argo profiles and their captured spatiotemporal correlations relevant for the AMOC reconstruction. In simpler terms, it compresses the complex ocean state into a format that the neural network can interpret and use to make accurate reconstructions. Argo profiles pose an important~~

Argo profiles pose a demanding challenge for the design of neural networks, as neural networks are normally designed to handle ~~fixed-size inputs. Sequence models like Long Short Term Memory (LSTM) cells (Lindemann et al., 2021) and Transformer based architectures (Wen et al., 2023) are an exception to this but still for the spatial component, a fix set of locations is required~~structured inputs. For spatio-temporal data points, structured input requires a constant number of measurements from constant locations with a static topology (e.g. a grid structure, or a graph), which is both not true for moving Argo floats. The virtual Argo profiles provide spatial information in form of longitude and latitude, and temporal information in form of a single point in time at which they reach the surface. ~~When neglecting the spatial information of the Argo profiles and interpreting the variable set of Argos as a sequence we can utilize a suitable neural network architecture, which will produce a fixed sized representation of the ocean state (Yan et al., 2022). Intuitively, this can be compared to modern text-AIs which process texts with variable input length and compute fixed-sized output summary, such an approach only covers a one-dimensional neighborhood with equidistant gaps between the sequence elements. Keeping the spatial component of the moving Argo profiles intact, however, requires a clustering approach to map the profiles into a structured space, but finding the right clustering is not trivial.~~ A static clustering on all data points~~would result in a zonal binning around~~, could create such a structured input. Around our targeted latitude ~~(?Hernández-Guerra et al., 2010)~~this would amount to a zonal binning (Willis, 2010; Hernández-Guerra et al., 2010) but the set of Argo floats, especially ~~on~~for the shorter target time scales, show a heterogeneous distribution that would ~~lead to information loss if a uniform cluster size is used. On the other hand, clustering each Argo set for each data sample would result in variable clustering and noisy assignments.~~ require carefully hand-crafted

cluster boundaries. A classical binning comes with even more assumptions about shape, distance, and connectivity of the bins, which would require a thorough testing of hyper parameters. However, we aim in this work for a data-driven mapping function which identifies a structured numerical representation (n-dimensional vector) of the underlying ocean state which is independent from specific spatial distribution of the Argo profiles. Often, the term embedding is used for such a n-dimensional vector, but also for the vector space, and for the mapping (Meilă and Zhang, 2024). Constructing a suitable embedding (see black box in Figure 2 B) has two main aspects. First, the same ocean state observed by different spatial and temporal distributions of Argo profiles should result in the same (i.e. negligible difference) embeddings. Second, two embeddings should be similar (i.e. small difference) if the corresponding ocean states are similar, too. A successful embedding creates a vector representation of the manifold of ocean states that can be used to generalize for unseen but similar ocean states in the evaluation.

~~For an architecture that combines the temporal part with spatially meaningful clustering, we utilized the framework~~

Therefore, we utilize a framework that learns a dynamic spatio-temporal clustering to create structured inputs. The framework Sparse Unstructured Spatio Temporal Reconstruction (SUSTeR, Wölker et al. (2023)) ~~which~~ was originally developed for road traffic reconstruction from sparse observations ~~. By randomly neglecting major parts of the traffic sensors (up to 99%) in the road network the data is made artificially sparse. Such sparse traffic sensors can be understood as they were moving in the road network and report spatially heterogeneous distributed measurements, similar to Argo profiles .~~ in varying locations. SUSTeR set out with the goal to handle unstructured traffic observations and find a general representation of city traffic, much like the unstructured virtual Argo profiles with the goal to find a general representations of the ocean state. For technical details we refer the reader to the original paper, and provide in the following a brief intuition for the oceanographic audience. This framework acts as a learnable clustering ~~in which an encoder adds each~~ with a fixed set of predefined clusters. The assignment decision is non-linear and adds each virtual Argo profile to one cluster based on ~~the training process. The assignment decision is based both on the spatial and the temporal component, where the~~ its spatial and temporal ~~component is expressed as the relative amount of days to the middle of the data sample (Learned Centroids in~~ component. A learnable clustering in contrast to the often used static clustering has dynamic boundaries between clusters which can represent different spatial and temporal resolutions, can have non-linear boundaries, or can change with global context (e.g. seasons). After each virtual Argo profile is assigned to one of the Learned Centroids (see Figure 2 ~~). Each Argo profile is assigned to a clustercenter, where its information is summed into the center's embedding. These cluster centers are then treated as nodes in a graph. Spatiotemporal correlations acting between different clusters can be learned by a graph neural networks which is a widely used approach in the field of traffic prediction (Jiang and Luo, 2022) and therefore have the potential to be a valid approach for Argo profiles. In contrast to the original SUSTeR architecture, which reconstructs traffic values for different traffic sensors, the goal of~~ B) the measurements of the virtual Argo profiles are aggregated within each cluster. These clusters represent the structured input, because their boundary will be fixed after training and the ~~AMOC reconstruction is to estimate a single value, the strength of the AMOC. Therefore, we do not use a decoder module that is queried for a given position, like the reconstructed traffic sensor location, but use the black box embedding as input in the feed-forward neural networkdecoder (Figure 2 C)), making the embedding in the black box a summary of the~~ amount of clusters is constant. To learn the spatial and temporal correlation between these dynamic clusters the

framework SUSTeR uses graph neural networks (Jiang and Luo, 2022), which can be thought of as message passing between the clusters, to learn the spatio-temporal correlations. After the clusters exchanged information, all features from the graph
385 nodes (learned clusters) are concatenated and reduced in their dimensionality by a shallow feedforward neural network. The result is not human interpretable but serves as an embedding representing the ocean state.

3.2 AMOC strength reconstructed with Neural Network

The output of the Argo processing module is an embedding which ~~will be an input to~~ is one input of the reconstruction module. ~~The content of this vector is defined by the training process of the neural network similar to the activations of a hidden layer.~~
390 ~~As both, the Argo processing and the reconstruction neural network are trained in an end-to-end architecture, this vector is in a black box and can not be explained.~~ We highlight the fact that the mean of all virtual Argo profiles in a single sample is subtracted from the profiles~~for the Argo processing and~~. The mean is separately used in the reconstruction module (see Figure 2 A), as the interior transport is characterized by a zonal pressure gradient~~which is~~, independent from the mean profile.

Together with the ten zonal wind stress bins, the Florida Current transport, the Antilles Current ~~tranport~~transport, and the
395 mooring information below the 2000m Argo depth, the data are concatenated into an input vector for a feed-forward neural network. A feed-forward neural network is a sequence of layers containing a linear mapping and a non-linear activation function. Such an activation layer is the Rectified Linear Unit (ReLU) which is a composite function of the two linear function ~~the~~ (constant zero for any negative ~~input and an identity function for values larger or equal than zero. By stacking such layers into a sequence, the architecture gets deep and increases the complexity of functions that it can approximate. The internal layers are
400 called hidden layers. The reconstruction module uses two hidden layers from an input of 365 dimensions containing the Argo mean profile in 29 dimensions one for each depth layer, 10 for the zonal bins of the wind stress, 1 for each current, 17 for the deep mooring information, and 12 dimensions for each of the 27 learned concepts from the Argo processing embedding, the first hidden layer maps this to 64 dimensions and the second to 32 dimensions. The number of learned concepts is set to 27 based on experiments that we evaluate for all timescales but that are not shown in this work. The last layer of the reconstruction
405 module is an output layer which is a linear mapping without any non-linearity to have an unbound value range for the output.~~ values and a identity function otherwise). Feed-forward neural networks represent the basic form of neural networks to find correlations in data that do not use information about the hidden neighboring structure of the data, i.e. if the data are to be understood as sequential ordering, as a regular structured image, or as a more flexible topology. Such networks were already used to reconstruct the AMOC in the work of Solodoch et al. (2023). We found that our architecture requires more neurons
410 compared to that of Solodoch et al. (2023) to handle the complex structure of the Argo processing vector representing the interior ocean state. The network is trained with backpropagation minimizing the mean squared error between the reconstruction and the ground truth ~~total or geostrophic AMOC~~interior geostrophic or AMOC transport, respectively. All input and output values are normalized with their respective mean and standard deviation for a faster training. ~~(Ioffe and Szegedy, 2015)~~ (Ioffe and Szegedy, 2015).

415 Further, we discuss the choice for the test, validation, and training periods of our data as it can be seen in Figure 1 e). While the data in the training period is used to iteratively train the reconstruction, the data in the validation period is used to

determine when to stop the training. All evaluations are only based on the data in test period. With the data source VIKING20X we decided to use 2004 - 2023 from the first OMIP cycle as the test period for the reconstruction because those results compare well with the RAPID time series and other key components of the AMOC in respect to the statistics and temporal evolution (Biastoch et al., 2021). As the task of the validation period is to hinder the leakage of training data into the test period, it should be adequately sized depending on the underlying temporal autocorrelations. We will investigate up to interannual signals, therefore we set aside five-year windows before and after the test period as validation data. While the earlier validation period 1999-2004 is trivial, the second validation period of five years starts with the second forcing cycle. But since the ocean state is continuing into the next cycle, we expect the later validation period to break the auto correlation of the interior ocean state in the same way the earlier period does.

3.3 Evaluation metrics

To evaluate the skill of the AMOC reconstruction in comparison to the VIKING20X AMOC, we use four different metrics, the mean absolute error (MAE), the root mean squared error (RMSE), the mean absolute percentage error (MAPE), and the coefficient of determination (R2). For a comprehensive overview we define those skill metrics in the following for a test period with N samples, y_i being the targeted AMOC, and $\hat{y}_i$ being the reconstruction.

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^{N} |y_i - \hat{y}_i| \tag{1}$$

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^{N} (y_i - \hat{y}_i)^2} \tag{2}$$

$$\text{MAPE} = \frac{1}{N} \sum_{i=1}^{N} \frac{|y_i - \hat{y}_i|}{y_i} \tag{3}$$

$$\text{R}^2 = 1 - \frac{\sum_{i=1}^{N} (y_i - \hat{y}_i)^2}{Var(y)} \frac{\sum_{i=1}^{N} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{N} (y_i - \overline{y})^2} \tag{4}$$

Large outliers are less visible in the MAE compared to the RMSE where errors are squared. The difference between MAE and RMSE gives an insight whether the reconstruction is able to keep a certain reconstruction quality throughout the test period. Both values are reported in the same unit as the original time series (Sverdrup ($1Sv = 10^6 \frac{m^3}{s}$)). Because the ~~total and geostrophic AMOC~~ interior geostrophic and AMOC transport have different amplitudes the ~~MAPE uses the MAE and divides it by the ground truth for a relative estimation of the error in the context of the ground truth making it possible to compare the reconstruction skill over different AMOC time series~~ relative MAPE allows for an comparison their reconstruction skills. The R2 score reports the amount of variability reconstructed from the target time series with 0% being a constant mean predictor

445 and 100% a perfect reconstruction. As we work with neural networks the skill is dependent on the initial random initialization of the weights~~because those will define the initial gradients in the training process.~~ In order to produce reliable results we execute each experiment eleven times ~~with different random initializations~~ and report the mean execution skills.

4 Results

4.1 Reconstruction of the AMOC at 26.5°N

450 To assess the reconstruction performance across a range of dynamical regimes, we evaluate AMOC reconstructions on timescales from 10 days to one year (Exp.1 in Table 1). Different drivers of the variability play a role on different time scales, which must be investigated to determine on which of the components, the reconstruction performs well and where it shows deficits. Short time scales are known to be driven by the atmospheric forcing in the study area, as shown by Kanzow et al. (2010), which enabled the AMOC reconstruction from Solodoch et al. (2023) to perform well on monthly data. They found a strong focus
455 of the reconstruction at the wind stress, but they kept the monthly reconstruction and low-pass filtered it to create longer time scales. Here, we show that an already low-pass filtered input will shift the focus onto the density structure, hence increasing importance of ~~Argo profiles~~the Argo component.
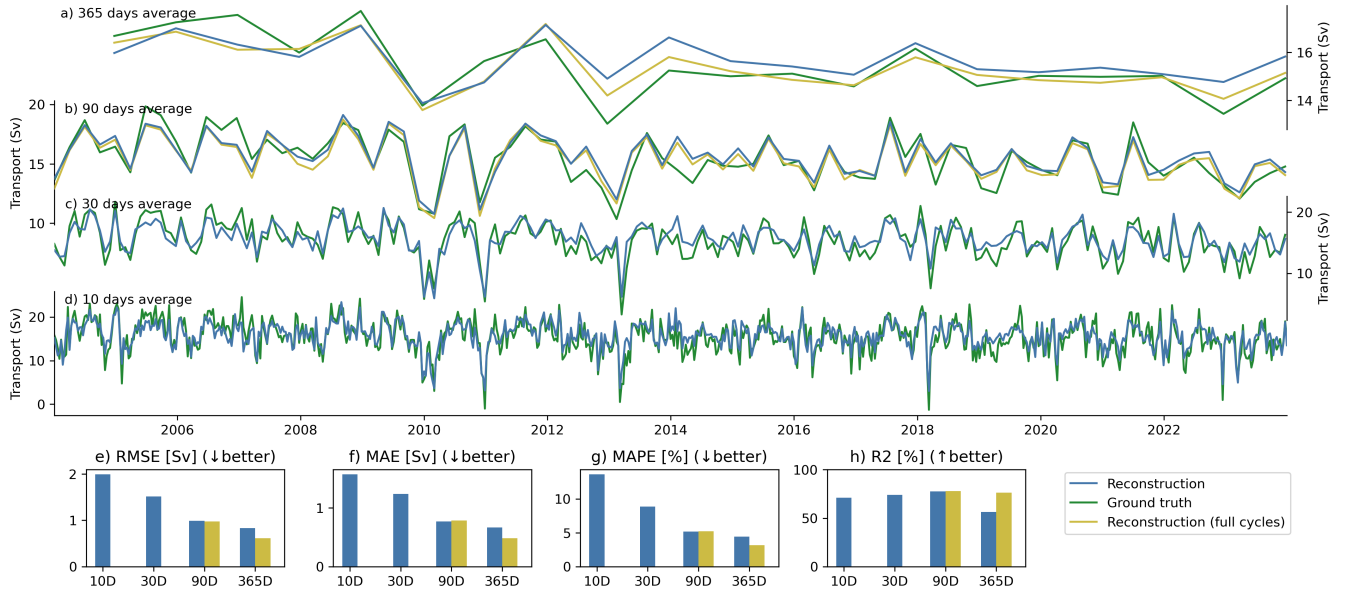
Figure 4. AMOC Reconstruction of the ~~total~~ VINING20X AMOC, trained on the first two cycles except for the test period from 2004 to 2023 of the first cycle and the validation of five years on each side of the test data for yearly a), seasonal b), monthly c), and 10-day timescales. The VIKING20X AMOC is depicted in orange, the reconstruction in blue, and the green reconstruction is trained with the additional cycles three to six. The skill of the reconstruction is measured in RMSE e), MAE f), coefficient of determination (R^2) g), and MAPE h).

The reconstruction generally captures AMOC mean and variability across all timescales (Figure 4), with decreasing accuracy on shorter time scales due to larger amplitudes. ~~The amplitude of the AMOC is the strongest for the short 10-day timescale~~

460 ~~(Figure 4 d)), which leads to the largest magnitudes of the value-based skill scores (MAE and RMSE) compared to other time scales. Overall smaller MAE or RMSE skill scores represent better reconstructions as both metrics measure the average difference of the target variable and the reconstruction. The the value-based skill scores are dependent on the magnitude of the time series limits the compare between different time scales, while the relative skill scores MAPE and R2 can be compared between those.~~

465 However, the short timescales not only show the largest mean absolute error (Figure 4 f)), also the largest relative error h) with over 10% mean error on average. We argue that those very short time scales exhibit ~~also~~ strong noise, which here we define as variability that is not represented in our input data. Examples for unexplained and therefore not reconstructed fluctuations on the 10-day time scales are changes in the spatial extent of the Antilles Current outside of our bounding box, a sparse amount of ~~Argos~~ Argo profiles that not cover the boundaries or other fine scale processes. For the latter, mesoscale

470 eddies and Rossby waves are potential signals that change the AMOC on such short timescales but are not resolved sufficiently in the input~~, although their signal can be part of the Argo profiles.~~This shows a limitation of the reconstruction approach for time scales of 10 days or shorter. Despite the limits of the reconstruction on short timescales, we highlight ~~the fact~~ that major features like the weak-AMOC events in end of 2010 or in the beginning of 2013 are well reconstructed (Biastoch et al., 2021). Short-term fluctuations are well captured, although the shortest 10-day time scales demonstrate a slight underestimation at the

475 beginning of the test period (2005-2007) and show reduced variability at a smaller scale in the subsequent years (2019-2021). Most of the skill is lost in the small variations which is the reason for a worse MAE and RMSE compared to a moderate skill decrease in the R2 compared to longer time scales.

Monthly reconstruction (Figure 4 c)) benefits from reduced high-frequency noise and shows a stronger signal-to-noise ratio. Compared to the 10-day timescale, all skill metrics improve for the monthly reconstruction. The signal-to-noise ratio improves

480 for the selected input variables used in the reconstruction, while we achieve an explanation of 74% of the overall variance (R2 score). ~~On average, the monthly reconstruction has an absolute error of 1.24 Sv, which corresponds to 8.8% of the targeted total VIKING20X AMOC.~~ It also shows robust performance, as is apparent from the standard deviation of MAE and RMSE across eleven runs with random seeds, which remains below 0.04Sv.

4.1.1 Extension to seasonal and yearly time scales

485 While the skill continues to improve on seasonal scales (Figure 4 b)), yearly reconstructions a) are constrained by limited training data. ~~Using only the first two OMIP cycles as training data, the yearly MAE, RMSE, and MAPE show no improvement in contrast to the seasonal time scale and the R2 score drops below 60% of explained variance. The largest deviation at yearly time scales (Figure 4 a)) between the reconstruction and the VIKING20X AMOC occurs between 2012 and 2017, where the reconstruction overestimates the transport. This offset is less apparent at shorter time scales.~~ It is known that neural networks

490 need vast amounts of training data to approximate the correlations in the data (Jiang and Luo, 2022; Ioffe and Szegedy, 2015) and exploit them for well performing reconstructions. At the yearly scale, each training sample corresponds to one year of
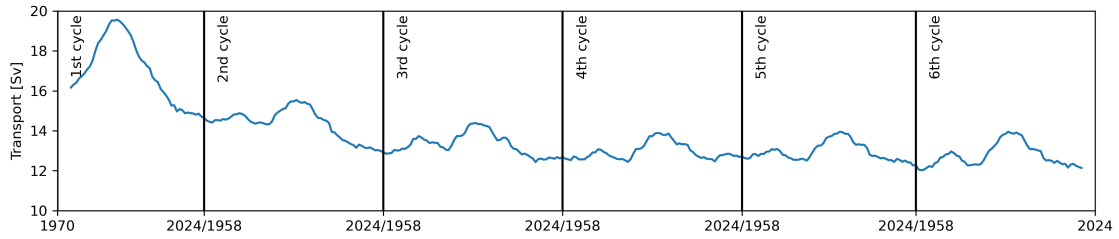
Figure 5. The AMOC strength for all consecutive six OMIP cycles of the ocean simulation VIKING20X. A 10-year low pass filter was applied to the timeseries. The vertical bars show the boundaries of the OMIP cycles and the x axis indicates the reoccurring forcing years.

data. Originally, we use years 1970–2024 from the first OMIP cycle and 1958–2024 from the second, which gives a total of 130 years ~~. After subtracting 20 years for testing and 5 years of validation on each side, only~~ (100 samples ~~remain for training.~~ ~~for training).~~ Considering that up to 2600 virtual Argo profiles appear in a single year within our bounding box around the target latitude only 100 data samples are a proportionate small number for a training process of neural networks. The original VIKING20X experiment contains six OMIP cycles with less AMOC variability in the later cycles (Schiller-Weiss et al., 2025). Although the six OMIP cycles reuse ~~wind stress patterns which are part of the reconstruction~~the atmospheric forcing, the additional cycles ~~still~~ provide valuable variability in the ocean state due to the long-term evolution of the AMOC and its link to the density structure and associated thermohaline circulation.

Extending the training dataset using all available OMIP cycles (Figure 4 ~~Reconstructin~~ Reconstruction (full cycles)) boosts performance on yearly time scales, and hence highlights the benefit of diverse training data for improving the neural network reconstructions. At the 90-day timescale, all four skill metrics show similar performance regardless of training data size, suggesting no improvement. However, the quantitative performance at the yearly time scale strongly benefits from the additional training data. ~~This highlights the importance for a heterogeneous ocean state dataset when neural networks are chosen as reconstruction method with the goal to work at unseen and independent time periods.~~ While the reconstruction for two cycle training and six cycle training mostly agree up to 2012 for the yearly timescale, the six cycle training is able to compensate the ~~overestimation of the reconstruction with only 100 training samples and agrees more with the ground truth from 2014 onward.~~former overestimation.

4.1.2 Generalization onto the third forcing cycle

A larger amount of training data helps with the yearly reconstruction skill but the variability on test data is limited due to the 20 year period. To show the reconstruction performance on a longer test period we choose the third OMIP cycle in Figure 6 as the test period~~and use~~ . Surrounded by 20 years of validation gap on each site~~to test the generalization of our approach on the longer time scales.~~.All other parameters of this experiment (see Table 1 Exp.2) are left unchanged compared to the previous~~experiment.~~ The third cycle shows a smaller AMOC amplitude across its two phases compared to the previously used first OMIP cylce. It has a strong phase during the forcing years from 1987 to 1997, and a weak phase from 2005 to

2015 (Schiller-Weiss et al., 2025), as seen in the ten-year low-pass filtered AMOC of all six OMIP cycles (Figure 5). While the third cycle is time-wise in the center of the training data this hold also for the AMOC strength in the respective phases. Since the AMOC experiences a long-term decline due to model adjustment, the earlier cycles have strong AMOC amplitudes while the subsequent cycles weaken even more than the third cycle, making it a suitable candidate to test generalization on intermediate conditions. This setup provides a longer test period for evaluating yearly reconstructions and includes both strong and weak-AMOC phases. It even allows for a preliminary assessment of reconstructions at five-yearly timescales.
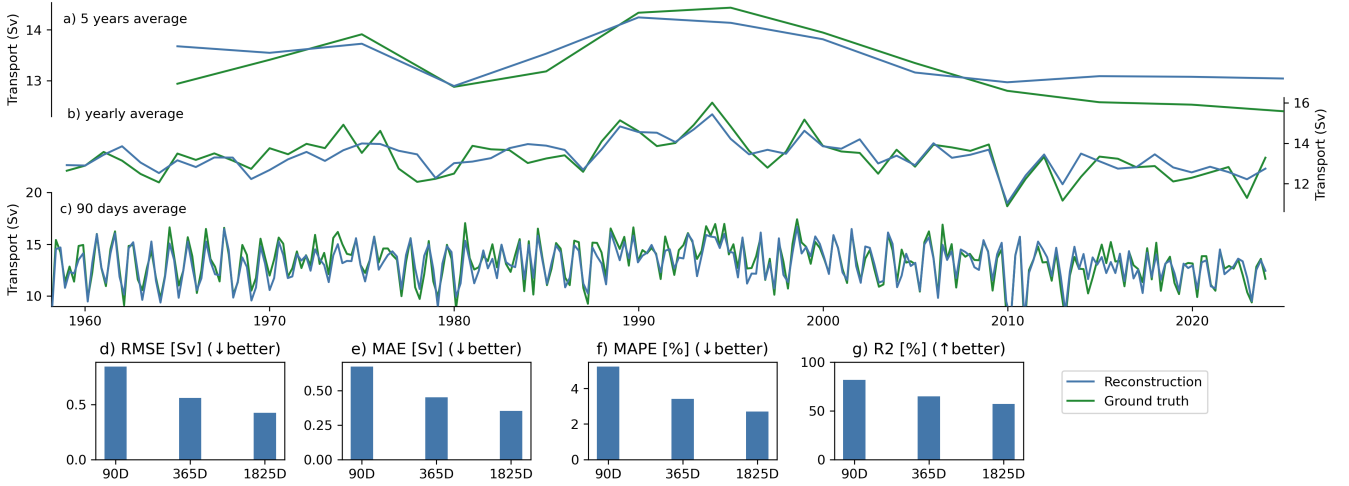
Figure 6. Reconstruction (blue) on five-yearly a), yearly b), and seasonal c) time scales with the AMOC of the third OMIP (orange) cycle as test period, validation period of 20 years were attached previous and after the test period, the rest of the six cycles was used for the training of the reconstruction.

The framework reconstructs interannual variability in the third cycle with similar accuracy to the shorter test period, capturing both strong-AMOC and weak-AMOC phases. At seasonal time scales (Figure 6 c), the architecture reconstructs the strong interannual signal across the full test period with similar performance (MAPE 5.2%; R2 81.8%) as in the two cycle experiment (Figure 4 a)). During the strong-AMOC years (1993–1997), the reconstruction shows lower interannual variability. This suggests that the model does not rely on a default cycle (e.g., surface temperature patterns in Argo) but adapts the amplitude to the specific conditions in that period. With the longer time series, the stronger AMOC phase is included in the test period and sets the dominant variance of the time series compared to the previous test period from 2004 - 2023 in the first cycle. Visually, the reconstruction mostly agrees with the ~~total~~ AMOC in the second half of the time series. From 1986 onward it matches the strong peak in 1995, and the low AMOC in 2010. However, on yearly time scales the reconstruction is slightly off in the 1970s and early 1980s, which was not observed on the seasonal and five yearly time scales. We speculate that this is connected to the lower quality of the wind forcing for the period before the SSM/I-based wind speed product start in 1988, which is used in the JRA forcing (Tsujino et al., 2018) that drives the VIKING20X simulation. Frajka-Williams (2015) showed that the interannual variance of the AMOC signal at the study latitude 26.5°N is driven by the zonal wind stress. The connection between the

535 forcing before SSM/I and the wind stress being the one driver of the interannual variance is our explanation for the misaligned reconstruction before the early 1980s.

As we evaluate a much longer test period we include also the five year time scale into this evaluation. Despite limited training data, the reconstruction shows promising skill on five-year averages (Figure 6 a)). It captures long-term AMOC shifts, although the reconstruction dampens the amplitude and can not reconstruct the full decadal variability. ~~More training data is~~
540 ~~needed to further examine and potentially reduce this overestimation. However, the same data scarcity issue applies as in the two cycle experiment with the yearly time scale. Nonetheless, the reconstruction is able to capture the strong-AMOC phase which reflects stronger deep-water formation in the subpolar North Atlantic (Biastoch et al., 2021; Böning et al., 2023) and the also the smaller AMOC peak in the mid-1970s which is missed in the yearly reconstruction. The reconstruction largely fails to represent the weaker-AMOC phase after 2005.~~ A major issue is the data scarcity but this experiment shows that
545 our reconstruction is likely to work ~~also~~ on longer time scales~~if more training data were available.Although for the yearly reconstruction .~~Although we used all six OMIP cycles, a sufficient number of five-yearly training samples would require additional OMIP cycles or additional ocean simulations, as we later discuss.

4.2 Importance of individual components for the AMOC reconstruction

To interpret the reconstruction process, we apply explainable AI (X-AI) techniques (Dwivedi et al., 2023) that reveal how each
550 input contributes to the neural network's estimation of the ~~total~~ AMOC. X-AI aims to make typical black-box behavior of neural networks transparent, providing insights into decision-finding process and evaluate the meaningfulness of the inference. The meridional transport at the RAPID array (26.5°N) represents an integrated sum of Ekman transport, transport in the Florida Strait, and upper-mid ocean geostrophic transport (Kanzow et al., 2008). For estimating the AMOC in the real ocean, Ekman transport can be estimated from satellite-based zonal wind stress, and Florida Current is available as a published product. In
555 contrast, the upper-mid ocean transport requires more effort, relying on moorings and repeated ship sections. We hypothesize that Argo profiles have a larger influence on the reconstruction for the longer time series. The main motivation here is that variability beyond the interannual timescales is mainly determined by thermohaline changes which are reflected in the density structure of the open ocean (Biastoch et al., 2008). Experiment 3 (see Table 1) calculates the feature importance (Fisher et al., 2019) for each input component (wind stress, Florida Current, Antilles Current, and Argo profiles) and will provide insight
560 in whether the reconstruction acts on longer time scales as expected. ~~To calculate the feature importance, the fully trained reconstruction neural networks as those from Figure 4 are used.~~ Feature importance quantifies how much each input component contributes to the reconstruction. The fully trained network from Experiment 1 (see Table 1) defines the baseline performance. By ~~deleting~~ removing the influence of an input component, the degradation of skill scores compared to the baseline describes the feature importance. As we always compare the degradation of the total performance, we achieve comparability among the
565 input components. To remove the influence of one input component, we randomly permute its values in the test data ~~. That breaks the correlation with all the other input components while preserving its data distribution, which would not be preserved, if the component was replaced with a constant value (e.g. 0). This random permutation is repeated 10 timesto create an average performance with the removed component and we report the average degradation in the following.~~ for ten times.
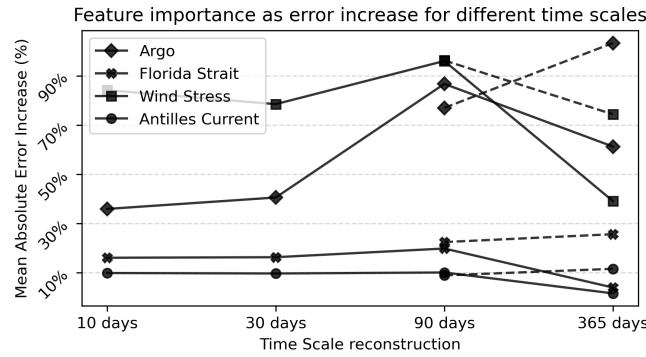
Figure 7. Feature importance for each input component based on the MAE ratio between the fully trained reconstruction baseline and evaluation with randomly permuted features. The larger the value is the higher the importance of this input component. The solid lines represent the two OMIP cycle trained reconstruction from Exp.1 (Table 1) and the dashed lines from the reconstruction using all six OMIP cyles.

The feature importance analysis reveals how different inputs influence the AMOC reconstruction across time scales, highlighting the growing role of Argo profiles with longer time scales. Figure 7 shows the feature importance with an higher importance being equivalent to a higher increase in the MAE skill when the component is removed from the evaluation. Across all time scales the importance of both currents Florida and Antilles are the lowest and stays constant around a 20% worse MAE skill score in contrast to the reconstruction with all input components. The black solid lines show the experiments with the two-cycle trained reconstruction two-cycles from Figure 4 while the dotted black lines represent the six-cycle trained version. Zonal wind stress and the Argo component show the highest importance with up 90% increase in the MAE. On short time scales the high importance of zonal wind stress is not surprising and agrees with current literature (Moat et al., 2020). However, a high importance of wind stress on shorter time scales partially validates the learned AMOC reconstruction, as it utilizes expected and known correlations.

Argo becomes even more important on longer time scales, while the wind stress importance stays constant or slightly decreases. For reconstructions trained on the first two OMIP cycles of VIKING20X (black lines), the importance of the Argo component increases up to the seasonal time scales. However, this pattern is also visible in the wind stress, which is in line with the previous finding that on shorter time scales the AMOC time series is more noisy leading to lower importance values. With both Argo and wind stress importance raising, still the wind stress is more important on time scales up to seasonal for the reconstruction. The turning point is the yearly time scale, which aligns to our hypothesis that on larger time scales, the Argo profiles take the lead from the wind stress for the most import important input component. While the importance of the wind stress strongly decreases and elevates the Argo profiles to the most important input component the gap However, the gap between these is marginal for the importance calculated only on two of the OMIP cycles. On the other hand, the feature importance calculated on the six OMIP cycle experiment (Figure 4) shows a more pronounced increasing trend of the Argo profile importance and a stronger change in the role of the most important input component on yearly scales from the wind

590 ~~stress to the Argo profiles.~~ In summary, reconstructions at longer time scales rely increasingly on the Argo component shown by an increasing focus of the AMOC reconstruction on interior measurements like Argo profiles.

In contrast to Argo profiles and wind stress, boundary currents contribute minimally to the AMOC reconstruction, especially on longer time scales. This finding is in-line with recent work about the signal-to-noise ratio (McCarthy et al., 2025), finding only a small signal in the Florida Current transport. However, feature importance only reflects the importance in

595 a fully trained reconstruction. It does not indicate if an input component can be left out from the input, referred to by fidelity (Markus et al., 2021). In VIKING20X, the Florida Current transport time series and the AMOC time series for the 90 days time scale have a linear correlation of 0.58 for the first two cycles shown in Figure 1. Removing the influence of the Florida Current transport increases the MAE by only 20%, suggesting that other inputs carry overlapping information. When calculating the feature importance with regard to the interior geostrophic transport, not shown in this manuscript, we

600 observed a similar pattern, the Florida Current gains importance for seasonal scales, due to the strong correlation to the AMOC (Frajka-Williams et al., 2016). Despite this the importance for all other inputs are constant for the interior geostrophic transport across all time scales with virtual Argo profiles being the most important.

4.3 Reconstruction of an RAPID-like ~~AMOC~~ interior geostrophic transport within the ocean model
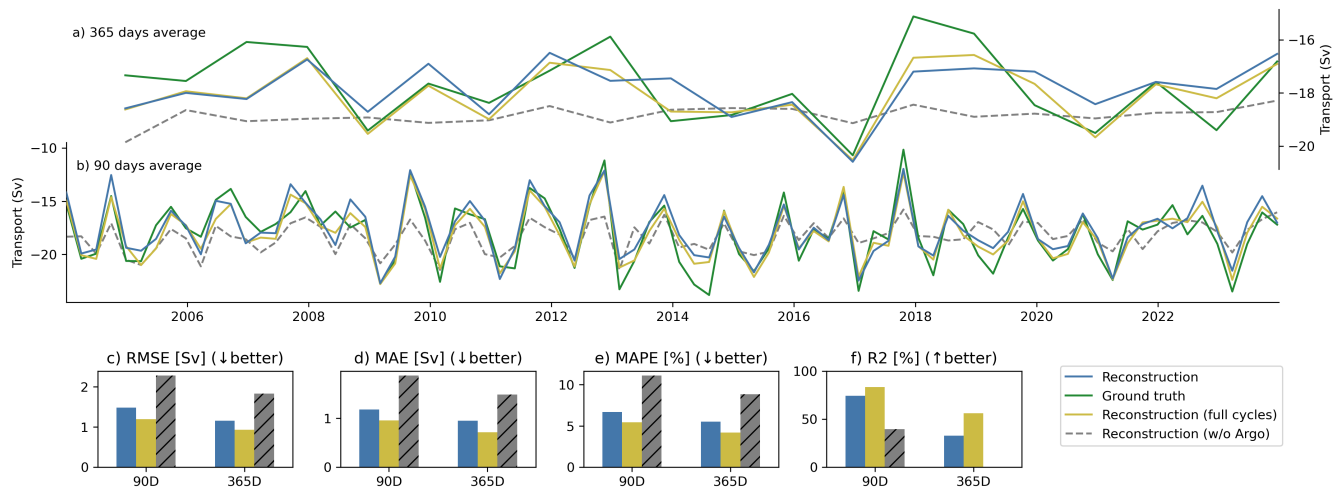
Figure 8. AMOC reconstruction trained with the RAPID-like interior geostrophic ~~AMOC~~ transport that is calculated from virtual moorings within VIKING20X in yearly a), and seasonal b) time scales. The skill metric of the time scales are shown in RMSE c), MAE d), MAPE e), and coefficient of determination R^2 f). The additional gray lines and bars show a reconstruction trained and tested without the Argo profiles but keeping all other input components.

This section investigates whether Argo profiles can help to effectively reconstruct the ~~upper mid-ocean geostrophic transport~~

605 ~~which currently relies on a complex and costly component~~ interior geostrophic transport of the RAPID ~~system.~~ upper mid-ocean ~~component.~~ ~~To align our calculations with the RAPID product, we isolate the interior transport from the total VIKING20X~~

~~AMOC by using virtual moorings (Figure 3). Note that all transports shown in this section, the RAPID-like geostrophic AMOC, mainly represents the southward deeper return brach of the AMOC, hence having a large southward (negative) transport.~~ The rising importance of Argo profiles at longer time scales (see Figure 7) suggests a stronger role of the geostrophic interior transport in AMOC variability. To validate our reconstruction further in this experiment (see Table 1 Exp.4), we define an additional target: a geostrophic interior transport time series (see section 2.4), aligned with the RAPID array's mid-ocean transport component. ~~The *total* AMOC in VIKING20X is calculated (see Section 2.4) as the sum over all grid box transports across the RAPID latitude 26.5°N which is an idealized setting not only due to the ocean simulation but also due the perfect coverage. Compared to the RAPID time series the mid-ocean component is computed from the Antilles Current, the interior transport from the mooring array and the residual mass transport(McCarthy et al., 2015). These components do not cover all parts of the section, but only those which can be constructed by rectangular areas between moorings. Hence, triangular regions on the slope and near-boundary portions of the full basin are unobserved by the moorings. Due to limited observations, the calculation of the mid-ocean time series relies on assumptions such as the applicability of an extrapolation towards the surface to compensate for the lack of near-surface measurement due to the swaying of moorings in the currents.~~

~~One goal of this study is to investigate the suitability of Argo profiles to reconstruct the interior transport in the context of measurement arrays like RAPID . To accurately represent this challenge, we do not use a grid-box-based geostrophic AMOC estimate but rather aim to apply the RAPID interior~~ The contribution of RAPID like interior geostrophic transport we find is to be predominantly negative transport. Between our calculated interior geostrophic transport ~~calculation by for example mimicking also the lack of full information which is present in the real ocean. Therefore, similar to the virtual Argo profiles, we extract virtual moorings from the ocean simulation VIKING20X, providing the foundation for our new ground truth. Each virtual mooring contains daily model output from the horizontal grid cell in which the original mooring is located and inherits the spatial and temporal distribution of missing values from the real-world RAPID mooring data. The sampling process from the 20 years of RAPID moorings onto the 60 years simulation cycles for the missing values is the same as for the position of the virtual moorings which is described in section 3.1. Using the virtual mooring data and the methods from McCarthy et al. (2015), we calculate the geostrophic transport between moorings, applying the same extrapolation to fill in missing data. We call this transport(Figure 3) our RAPID-like *geostrophic* AMOC which mimics the interior mooring-based transport estimates in the RAPID mid-ocean timeseries. Between our geostrophic AMOC~~ time series and the ~~total~~ AMOC we found a linear correlation of 0.38 on the seasonal timescale showing that the limitation of observations and missing boundary and surface processes change the characteristic of the geostrophic time series. Nonetheless, targeting a reconstruction of this geostrophic time series will provide insight in the suitability of ~~Argos~~ Argo floats to be a proxy for the interior transport ~~that is covered by the moorings.~~ as it is measured by arrays like RAPID.

We now evaluate how well the reconstruction framework can capture the geostrophic component of the AMOC. The geostrophic reconstruction uses all input components to utilize all possible data correlations, which is guided by the fact that the different components are not completely independent. For example, it was found by Frajka-Williams et al. (2016) that the Ekman signal contains information about the deeper density structure and the western boundary current. The ~~geostrophic AMOC reconstruction~~ reconstruction of the interior geostrophic transport (see Figure 8) performs best at seasonal time scales,

revealing differences in predictability between AMOC components. On time scales shorter than the seasonal scale, the reconstruction is not comparable to the performance of the ~~total~~ AMOC which is expected because the variability of the ~~total~~ AMOC is mainly wind-driven (Kanzow et al., 2010). Hence, we not include those shorter time scales in the evaluation.

At seasonal resolution (see Figure 8 b)), the reconstruction captures prominent interannual features of the ~~geostrophic AMOC~~ interior geostrophic transport, suggesting a physically meaningful signal. The time periods with strongly reduced interannual variability in 2013, and 2022 are well reconstructed, suggesting that our reconstruction framework does not compute a static seasonal signal but rather adapts to the dynamical regimes. Although we do not use any absolute timing feature as input (e.g. year, month, or day) from which such a periodical signal could be trivially inferred. However, we believe that information about the phase of the seasonal cycle is still present in the hydrographic data and hence, could still be inferred by the reconstruction. For example, Argo profiles include mixed-layer measurements, which in the subtropical North Atlantic are sufficient to infer the seasonal cycle from surface-near layers. Mostly the seasonal reconstruction struggles with the strong southward events as in 2013, and 2014 by underestimating the amplitude.

Despite improvements at seasonal scales, the reconstruction of the ~~geostrophic AMOC~~ interior geostrophic transport loses accuracy at yearly resolution, pointing to limits in data coverage and training diversity. With the yearly time scale we see a similar drop from seasonal to yearly in the explained variance for the two cycle OMIP training of the *total* AMOC (Figure 4 g). Interestingly, the reconstruction trained on all OMIP cycles with the geostrophic target variable drops to 50% of explained variance (Figure 8 f)) which in contrast kept stable for the *total* AMOC. The additional cycles from the two cycle training to the six cycle training in the geostrophic reconstruction increase the performance on the seasonal time scale across all metrics. This improvement was not observed with the ~~total~~ AMOC, indicating that more training data benefits the geostrophic time series reconstruction similar to the more training data for the yearly reconstruction of the ~~total~~ AMOC. This opens the question whether the yearly reconstruction of the ~~geostrophic AMOC~~ interior geostrophic transport can still be improved by more heterogeneous ocean states in the training data. One option is to reuse the same training samples but using different spatial Argo distributions to augment the existing data; another is adding additional ocean simulations to the reconstructions. However, both approaches begin to address the broader problem of transfer learning for applying reconstructions to real-world data, which we outline in the discussion but lies beyond the scope of this study.

Important for our study is the comparison of the ~~geostrophic~~ interior geostrophic transport reconstruction with the Argo profiles as described before and a reconstruction trained without any Argo profiles or information about the deep ocean. Such a reconstruction that only utilized zonal wind stress, Florida Current transport, and Antilles Current transport is shown in the gray dashed lines in Figure 8 a,b) and as gray bar in the corresponding skill plots c-f). We retrained the reconstruction in a neural network without argo-related inputs. This removes the influences of the Argo profiles completely from the reconstruction. Across all skill score a strong performance drop is present between the reconstruction with and without Argo profiles, highlighting the importance of ~~depth information~~ interior measurements for our reconstruction ~~even on the two cycle training~~. While the seasonal comparison shows that the reconstruction without Argo profiles underestimates the amplitude of the seasonal variability, the reconstruction on the yearly scales can hardly reconstruct any interannual variability but rather predicts a constant mean function. Despite the limited training data for the ~~geostrophic AMOC~~ interior geostrophic transport at yearly

time scales the comparison to an reconstruction without any Argo floats pinpoints the importance of the Argo component in our reconstruction.

4.4 Importance of the deep geostrophy
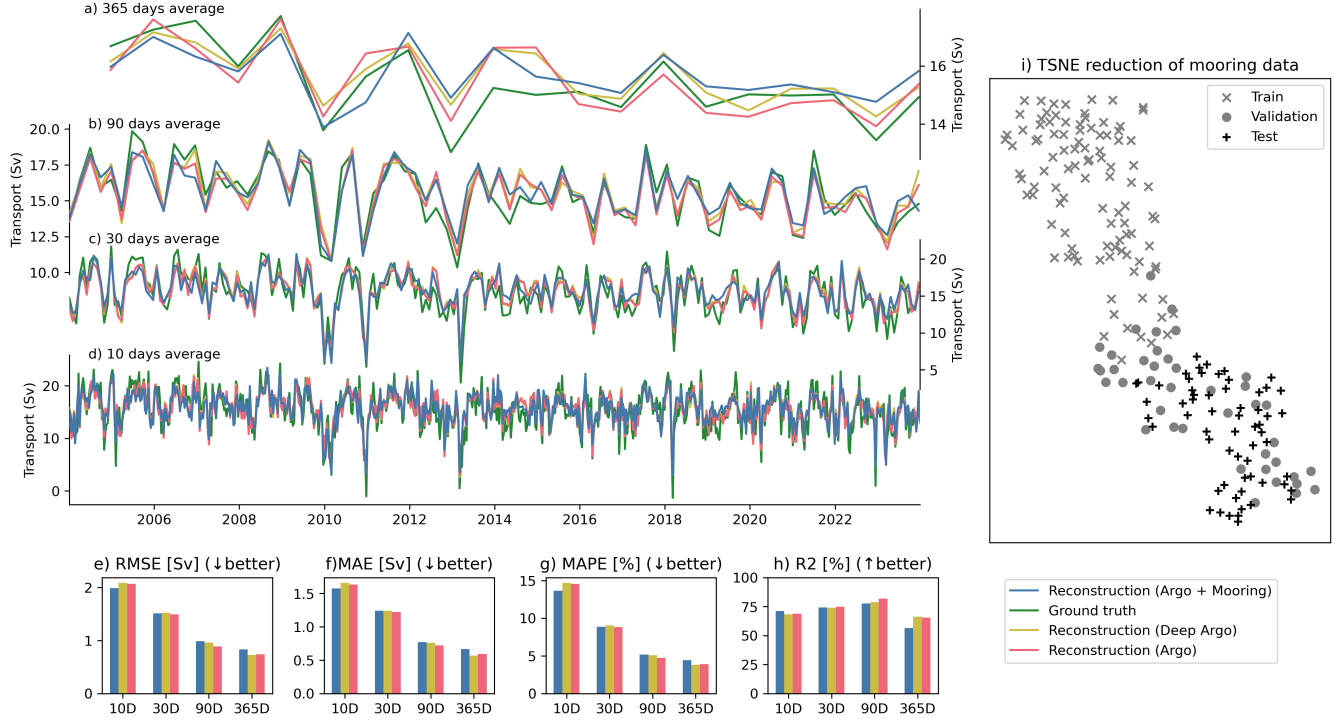
Figure 9. Reconstructing the ~~total~~ VIKING20X AMOC on yearly a), seasonal b), monthly c), and ten-days d) time scales with Argo reaching 2000 meter depth (red), Deep ~~Argos~~ Argo floats reaching up to 6000 meter depth (yellow) and ~~Argos~~ Argo floats up to 2000 meters as well as the deeper mooring information (blue). The skill is shown in RMSE e), MAE f), MAPE g), and coefficient of determination R^2 h). For a deeper insight into the performance we show the TSNE reduction of the ocean state estimation (Figure 2 B)) for the ~~seasonaltime~~ seasonal time scales and the mooring version split into training, validation, and test data.

680   The standard Argo network, which samples the ocean only down to 2000m, may not capture key features of deep ocean variability which could influence AMOC reconstructions. Thus, the reconstruction cannot rely on information from layers below ~~2000 m~~2000m, raising the question of how essential deeper data is for accurate reconstruction. In the previous experiments, the reconstruction was supplemented with the virtual moorings between the bottom and 2000m depth. ~~Other options to have an insight into the depth are Deep Argos (Roemmich et al., 2019), which reach depths of 6000m and have recently begun to~~

685   ~~grow in number.~~ To assess the potential value of deep-ocean information, we compare in this experiment (see Table 1 Exp. 5) reconstructions using standard Argo profiles, exchanging all Argo profiles with Deep Argo profiles (to 6000m), and standard Argo profiles supplemented with mooring-derived transport below 2000m. In the Deep Argo experiment, we assume a hypo-

thetical scenario in which Deep Argo floats offer the same spatial coverage as as the standard Argo array. ~~Hence, we repeat our experiment with virtual Argo profiles extracted up to 6000m depth and discard the supplementary mooring information~~

690 ~~formerly provided below 2000m, shown in Figure 9 in green. As a reference for this Deep Argo experiment, we add an experiment which only relies on 2000m Argo profiles (red lines) and does not use the virtual mooring data for supplementing in the deep to isolate the skill effect of the mooring data.~~

~~Across all timescales, reconstruction performance is surprisingly similar regardless of depth coverage, suggesting limited added value from information below 2000m except in specific contexts~~Across all timescales, we found that the added value

695 is limited by the seen training data, which did not cover heterogeneous ocean states in the deeper layers, leaving the possible influence of Deep Argo floats undetermined in our training setup. The similar patterns in the reconstruction suggest that information below 2000m may have limited influence on the reconstruction performance, even though minor differences appear in the skill metrics (MAE, RMSE, and MAPE). The virtual standard Argo profiles combined with the virtual moorings perform best at the 10-days time scale, but show the worst performance at longer time scales. This is even more highlighted in the R2

700 score in Figure 9 h), which shows the expected decline in performance at yearly scales. Interestingly, the decay is the strongest for the combined Argo and mooring data.

To understand the performance drop in the mooring-supported reconstruction, we use t-distributed Stochastic Neighbor Embedding (t-SNE (van der Maaten and Hinton, 2008)) to assess whether deep ocean states in the test set are represented in the training data. The algorithm is part of the manifold learning family and maps points from a high-dimensional space into

705 lower dimensions ~~for an better visualization . The mapping is constrained to keep as much from the neighboring structure as possible. Each virtual mooring sample is a high-dimensional vector, where each depth layer is one dimension.~~ visualization purposes. Using t-SNE, we visualize clusters of the virtual moorings (below 2000 meter) across the training, validation, as test data. ~~We test if the test data lays within the training data and its inference is an interpolation of seen training data.~~ As shown in Figure 9 i), the test data is almost completely separated from the training data, indicating a significant distribution shift between

710 them.

These findings suggest that while deeper observations can improve reconstructions, their benefit depends on having sufficient representation in the training data. The performance drop occurs because, during the test period, the reconstruction is forced to extrapolate in unseen data regions. To better leverage mooring data, the training must include more heterogeneous deep ocean states that are representative for the test period.

715 **5 Discussion**

In this study, we presented a framework that is based on neural networks to reconstruct the Atlantic Meridional Overturning Circulation (AMOC) across time scales ranging from 10-days to yearly and longer using observation-like input data. We trained the framework on consecutive forcing cycles produced under the OMIP protocol within the eddy-rich VIKING20X ocean model, focusing on both the ~~total~~ AMOC and the geostrophic component. The latter mimics the interior transport from

720 the AMOC, estimated based on the RAPID-array moorings. Our results show that a) reconstruction skill varies significantly

with time scale and input type, b) Argo profiles are a promising replacement option for the reconstruction of the RAPID-like interior geostrophic transport, and c) the amount and heterogeneity of the training data is important especially for accurately capturing the longer time scale variability. A core strength of the framework is its use of a graph neural network that is tailored to sparse, and irregular distributions of Argo floats, enabling the model to learn spatio-temporal patterns from realistic Argo float distributions without the need for interpolation or binning to spatially regular structures. These findings underline the potential of data-driven reconstructions with observation-based measurements for variability estimates of the AMOC and highlight the data limitations that must be addressed to enable reliable long-term AMOC monitoring with Argo floats.

In accordance with previous studies (Solodoch et al., 2023), we also found a strong relation between the zonal wind stress and the AMOC on short time scale of up to 30 days. However, we can not reproduce the results of over 90% explained variance seen by Solodoch et al. (2023). This might have different reasons. We have seen that the AMOC undergoes substantial variability on pentadal and decadal time scales which can be understood as a changing background state of the AMOC. Testing a reconstruction on previously unseen decades of data requires generalization of the reconstruction to different decadal background states, a challenge which is not present for the 5-year testing period available to ~~(Solodoch et al., 2023)~~ Solodoch et al. (2023) and the absence of a sufficient validation period that isolates training from testing data. Our 20-year testing period padded with 5 years of validation data ~~on the other hand~~ requires our reconstruction to handle the underlying decadal variability. Hence, it is not surprising that our explained variance does not meet the very high values seen in the previous study. Due the length of our test period and the heterogeneous states of the AMOC which are covered in each of the OMIP cycles we can not apply a linear detrending algorithm to ~~our time~~ the transport series as previous reconstruction studies did (Sanchez-Franks et al., 2021; Solodoch et al., 2023). This highlights that for achieving a reconstruction that potentially generalizes to unseen dynamical regimes, the training and test data need to be chosen and preprocessed carefully.

On longer time scales, we identified an increasing importance of the Argo profiles which we used as a widely available observation product to monitor the upper cell of the Northern Atlantic Deep Water at the RAPID latitude. While other studies use ocean bottom pressure (OBP) for the estimation of the interior ocean transport (Solodoch et al., 2023), we highlight the large effort for calibration and the instrument drift (Herrford et al., 2021)~~which could mean that an approach based on OBP instruments may not be suitable for long-term monitoring. The satellite-based~~ .Satellite-based OBP estimates from e.g., the GRACE mission (Landerer et al., 2020), require a high spatial resolution for a good reconstruction at the continental shelf ~~around the study latitude which currentlty is hard to achieve~~ (Delman and Landerer, 2022). Hence, satellite-based OBP estimates may be promising for ~~reconstructing the AMOC~~ AMOC reconstruction in the far future when decades of recorded data exists from satellites yet to be deployed. The advantage of Argo floats over satellite-based OBP is their longer availability since the early 2000s. The advantage of Argo floats over mooring-based estimates is their presence in regions currently not covered by moorings. Hence, and although other regions may have lower coverage by Argo floats, there is potential to train an AMOC reconstruction ~~based on Argos float data to be useful~~ at other latitudes. ~~Argo floats are only available in the last 20 years but because we used 120 years in our training data, we had to sample from the distribution of Argo profiles. We did so by mimicking for each simulation data point from the 120 years a realistic Argo distribution by taking a random but consecutive time period from the 20 years of available real-world Argo positions. Consecutive dives of a single float are not~~

~~linearly independent and therefore have drastically reduced degrees of freedom in the distribution compared to just sampling randomly distributed profiles. Accurately capturing the overlapping information from consecutive dives of Argo floats is a necessary condition for possible later transfer of the framework trained on ocean models to real-world measurements.~~ Using

760 Argo profiles showed a great performance to approximate the geostrophic component of the AMOC with the major limitation being the amount of available training data. We have shown that all six OMIP cycles have to be used in the training for a reconstruction on yearly time scales~~which we explain by the large amount of Argo profiles that are measured in a single year and the therefore larger effort for the reconstruction to approximate the spatio-temporal correlations.~~ _._ Interestingly, the additional later OMIP cycles (3-6) increased the reconstruction performance for both the ~~total and geostrophic AMOC although~~ _AMOC and interior geostrophic transport. Although,_ the strength and variability of the later cycles is lower compared to the

765 test period in the first cycle. Such an improvement shows that additional training data ~~could~~ _can_ be acquired from data that represent other ocean states, making a case for training with a model ensemble for potential further improvement. However, data from different OMIP cycles do not represent linearly independent samples due to the repeated atmospheric forcing~~and its associated seasonal and interannual variability.Hence~~ _.Hence,_ using over 300 years of data from all OMIP cycles is not comparable to a 300-year coupled ocean atmosphere simulation~~, where the available ocean states would be more independent~~

770 ~~and also cover more atmospheric states.~~ _._ A particular advantage here is the long-term, model-related, drift of the AMOC in VIKING20X (Schiller-Weiss et al., 2025). We utilized this to verify the performance of the Argo profiles ~~also on a different test period of~~ _at_ the complete third OMIP cycle ~~with a validation period on each site of 20 years~~ (Exp. 2 Table 1). It showed a well-reconstructed annual-mean AMOC and highlighted that our approach is applicable in other _transient_ dynamical states of the AMOC~~, such as that of the third cycles where the AMOC transitions from large-amplitude first two cycles to the less~~

775 ~~variable last cycles.~~ _._

  We isolated in this study the interior transport in ~~an~~ _a_ RAPID-like fashion from the ocean model output. Therefore we utilized virtual moorings to best approximate the method of the RAPID upper mid-ocean transport component with it coverage rate and assumptions. We focused on the geostrophic part of the upper mid-ocean RAPID component to investigate to which extent Argo profiles can be helpful in reconstructing the deeper southward return flow of the AMOC that is based on ~~geostrophy~~

780 _geostrophic transport_ (Elipot et al., 2014). For longer time scales, we found a good skill of reconstructing the geostrophic part of the AMOC from Argo floats. This is a promising finding for possible later transfer of our approach to real-world observations. From the RAPID array, the mooring array covering the interior transport is the perfect test bed for assessing to which extent neural networks can estimate the interior transport based on Argo profiles. ~~In this study we~~ _We_ delivered a first proof of concept by calculating a RAPID-like interior transport and achieving a well-reconstructed ~~virtual geostrophic AMOC~~

785 _time series at 1000m depth_ based on virtual Argo data.

  For a routine application on the RAPID array and other arrays, this study also identified limitations. The training within VIKING20X with _virtual_ Argo profiles has shown ~~the requirement to have large amounts diverse set of~~ _that an AMOC reconstruction on interannual or longer time scales requires large amounts of diverse_ training data. This limits a naive application of our framework to real Argo floats due to only 20 years of available data with a constant coverage of Argo profiles

790 only after 2012. ~~Transfer-learning~~ _However, for shorter timescales and use cases like the filling of smaller temporal gaps the_

observational data within the next decade could be sufficient to train an AMOC reconstruction purely on real-world data. A promising approach for longer time scales are transfer-learning methods, which allow neural networks to be pre-trained on larger similar datasets, e.g. from model simulations, that are then ~~transformed~~ transferred and executed on scarce target-domain data, e.g. real RAPID observations~~, are needed to adjust a learned reconstruction trained on ocean simulations to work on real~~

795   ~~observations. .~~ Such a transfer performs best, the more similar the data of the training and the targeted domain are. Due to model biases it is not guaranteed that ocean simulations have realistic spatial distribution of heat and salt, although the variability of the large scale measurements like AMOC appear to match. A further reduction of those model biases and an even higher spatial and temporal resolution will certainty be helpful for a transfer albeit possibly not feasible in the near future. ~~Although longer~~ Longer training periods are already seen in literature for the reconstruction of AMOC with neural networks as for example

800   8000 annual means from the Community Earth System Model presented by (Wu et al., 2025). However, we argue that for a future application to the real-world ocean, longer simulations of high-resolution ocean models ~~are necessary and will increase the chance of successful transfer learning from simulation based training to inference in the real world. Another approach is the usage of a model ensemble with different characteristics and extents of water masses like the CMIP comparison (Heuzé, 2021) where each ensemble member introduces heterogeneity to the reconstruction and hence diversifies the learned oceans states~~

805   ~~as well as the atmospheric states due to the coupling of both components. Similar to the reason that more trust is assigned to the mean of an ensemble this would be the motivation to learn a transferable reconstruction from a heterogeneous ensemble. Owing to the importance of resolution to represent the key components of the AMOC such as subpolar deep-water formation and overflow, a high-resolution version of the CMIP effort would be an interesting starting point for such a direction.~~ will be necessary to minimize the bias.

810   Nonetheless, our study points out that Argo profiles can be a promising alternative for the reconstruction of interior transports together with the finding that Deep ~~Argos and Mooring~~ Argo floats and mooring data below 2000m can not yet be used for our reconstruction approach. This is contrary to the large signal to noise (SNR) ratio that McCarthy et al. (2025) found between the Lower North Atlantic Deep Water and the AMOC. But the SNR based on a linear regression is calculated with five year filters, suggesting that the deep geostrophy will play a role on timescales longer than those investigated in this study.

815   The reconstruction of the AMOC time series from related observational instruments (e.g. satellite-based altimeters for measuring sea-surface height) is of large interest to monitor its variability at other latitudes or reduce the cost for measurement arrays like RAPID, or OSNAP. Such reconstructions can be based on various methods ranging from purely statistical (Willis, 2010; Frajka-Williams, 2015; Sanchez-Franks et al., 2021) to very powerful AI methods (Solodoch et al., 2023; Wu et al., 2025). Our study fits with this theme showing the potential and opening an interesting direction of research with basin-wide

820   Argo profiles and advanced AI methods to reconstruct the AMOC strength in a high resolution ocean model. However, as mentioned, the transfer from the ocean model to a reconstruction of the real-world RAPID measurements will require further investigations. Here, we set the stage for these further investigations by reconstructing a RAPID-like ~~geostrophic AMOC~~ interior geostrophic transport and by elaborating on the importance of training data sizes and the potential leakage from training to the evaluation. This work emphasizes the chance for model-data fusion using advanced AI methods to better monitor the

825   real-world AMOC at different latitudes of the Atlantic Ocean.

Code and data availability. The artifacts of this study are divided into two datasets and two software artifacts. The study is based on an isolated software framework(Wölker, 2025a) to ensure further development of the approach. For the reproducibility of our results and the approach we published the input components and target variables that we extracted from the ocean simulation in this dataset (Wölker et al., 2025). To rerun the experiments or recreate the figures of this study we published another software repository (Wölker, 2025b). In addition, we published the experimental results for this study which are the base for the figure creation (Wölker, 2025c). The Argo data were collected and made freely available by the International Argo Program and the national programs that contribute to it. (https://argo.ucsd.edu, https://www.ocean-ops.org https://doi.org/10.17882/42182#110199). The Argo Program is part of the Global Ocean Observing System.

Appendix A: RAPID-like interior geostrophic transport calculation

To create a interior geostrophic transport time series, comparable to the interior transport estimates from the RAPID array, we implemented our own calculation from virtual moorings in the VIKING20X ocean model. Here, we briefly describe our assumptions and the calculation from the daily-mean ocean model output. The overall process is based on the description from McCarthy et al. (2015) and has two main steps, i) the extraction and merging of mooring-based density profiles, ii) calculation of the basin-wide transport. All steps are also part of the published software code (Wölker, 2025a).

Extraction of virtual mooring density profiles

For a given mooring location we identify the closest location on the $1/20°$ model, calculate in-situ density for each z-level and linearly interpolate the resulting profiles to 20 meter z-levels, corresponding to the published RAPID mooring data. We extract virtual mooring density profiles for the following RAPID mooring codes: WB3, WBH2, WB2, MARWEST, MAREAST, EB1, EBH1, EBH2, EBH3, EBH4.

We group the virtual in-situ-density profiles into four groups representing the western and eastern boundary and the western and eastern flank of the Mid-Atlantic Ridge. Then we merge the profiles in such a way that for each z-level, the resulting profiles always contain the density estimate that is closest to topography. This results in four virtual in-situ-density mooring profiles, the two boundary moorings (ρ_W, ρ_E), and the two moorings (ρ_{MW}, ρ_{ME}) at either side of the Mid Atlantic Ridge.

Calculation of basin-wide geostrophic transport

The goal is to estimate a transport time series for the $1000m$ z-level. To this end, we first calculate vertical shear for three compartments of the zonal section. Two deeper compartments span the waters between the ocean floor and $3700m$ depth west and east of the Mid-Atlantic Ridge. The third compartment spans the water body between the eastern and western boundary between $3700m$ depth and the surface. As the real RAPID mooring profiles do not cover the water column all the way up to the surface, and as we want to mimic the data coverage of the real RAPID moorings as closely as possible, we mask out virtual-mooring data shallower than a certain depth z_{mask} which is different for each real RAPID mooring profile we are mimicking. Therefore, we sample uniformly from the corresponding RAPID mooring the depth shallowest instrument depth, similarly as we sample also the position of the virtual Argo profiles. We compute the zonal mean of vertical shear of northward

velocity for all z-levels z_i deeper than z_{mask} from the virtual density profiles and assume constant shear for depths shallower than z_{mask}.

$$\frac{\partial v}{\partial z_i} = \begin{cases} -\frac{g}{\rho_0 f} \frac{\rho^W_{z=z_{mask}} - \rho^E_{z=z_{mask}}}{x^W - x^E} & \text{for} \quad z_i > z_{mask} \\ -\frac{g}{\rho_0 f} \frac{\rho^W_{z_i} - \rho^E_{z_i}}{x^W - x^E} & \text{for} \quad z_{mask} \geq z_i \geq -3700m \\ -\frac{g}{\rho_0 f} \left(\frac{\rho^W_{z_i} - \rho^{MW}_{z_i}}{x^W - x^{MW}} + \frac{\rho^{ME}_{z_i} - \rho^E_{z_i}}{x^{ME} - x^E} \right) & \text{for} \quad -3700m \geq z_i \end{cases} \tag{A1}$$

860 For later use, we define the width of the water body at depth z as $L_x(z) = x^W - x^E$ for the upper compartment and $L_x(z) = x^W - x^{MW} +$ for the lower compartment. From, this, we can estimate zonal mean northward velocity at depth z referenced to the surface velocity as

$$v(z)|_{0m} = \int_{0m}^{z} \frac{\partial v}{\partial z_i} \, \mathrm{d}z' \tag{A2}$$

and northward transport relative to a the surface as

865 $$V(z)|_{0m} = \int_{0m}^{z} L_x(z') v(z')|_{0m} \, \mathrm{d}z' \tag{A3}$$

Following McCarthy et al. (2015) we choose a reference level of no transport of $4820m$ marking the bottom layer of RAPID measurements and not including the overturning of the Antarctic Bottom Water. Then we calculate the interior geostrophic transport IGT contributing to the $1000m$ overturning as

$$IGT = V(-1000m)|_{-4800m} = V(-1000m)|_{0m} - V(-1000m)|_{-4820m} \tag{A4}$$

870 Selecting the $1000m$ overturning avoids the detection Antarctic Intermediate Water (AAIW) as done by (McCarthy et al., 2015) . However, in a direct comparison to the Rapid timeseries two reconstructions can be trained one for 1000 meter and another for 700 meter. We argue that such an extraction following RAPID as close as possible is necessary to create an accurate time series that contains all the characteristics and caveats of the observations and to train the AMOC reconstruction most realistically to find reasonable correlations. A similar computed AMOC series is a strong prerequisite for any assumptions
875 about transferability between the ocean model VIKING20X and the real-world observations within RAPID.

Author contributions. AB: initiated the idea, conceptualized the approach, analysis of the results; MR: initiated the idea, design of the machine learning method; WR: contributed to design of technical concepts, improving the validity of the approach, analysis of the results; YW: implemented and performed the experiments, design and implementation of the approach and machine learning method, analysis of the results, written the first draft - All authors reviewed and improved the draft over iterations.

880   *Competing interests.*   The authors declare that they have no conflict of interest.

Acknowledgements.   The study was supported by the European Union's Horizon 2020 research and innovation program under Grant Agreement 818123 (iAtlantic). The authors gratefully acknowledge the Earth System Modelling Project (ESM) for funding this work by providing computing time on the ESM partition of the supercomputer JUWELS at the Juelich Supercomputing center (JSC).

References

885   Baker, J. A., Bell, M. J., Jackson, L. C., Vallis, G. K., Watson, A. J., and Wood, R. A.: Continued Atlantic overturning circulation even under climate extremes, Nature, 638, 987–994, https://doi.org/10.1038/s41586-024-08544-0, 2025.

Behrens, E., Biastoch, A., and Böning, C. W.: Spurious AMOC trends in global ocean sea-ice models related to subarctic freshwater forcing, Ocean Modelling, 69, 39–49, https://doi.org/10.1016/j.ocemod.2013.05.004, 2013.

Biastoch, A., Böning, C. W., Getzlaff, J., Molines, J.-M., and Madec, G.: Causes of Interannual–Decadal Variability in the Meridional Overturning Circulation of the Midlatitude North Atlantic Ocean, Journal of Climate, 21, 6599–6615, https://doi.org/10.1175/2008JCLI2404.1,

890   2008.

Biastoch, A., Schwarzkopf, F. U., Getzlaff, K., Rühs, S., Martin, T., Scheinert, M., Schulzki, T., Handmann, P., Hummels, R., and Böning, C. W.: Regional imprints of changes in the Atlantic Meridional Overturning Circulation in the eddy-rich ocean model VIKING20X, Ocean Science, 17, 1177–1211, https://doi.org/10.5194/os-17-1177-2021, 2021.

895   Böning, C. W., Wagner, P., Handmann, P., Schwarzkopf, F. U., Getzlaff, K., and Biastoch, A.: Decadal changes in Atlantic overturning due to the excessive 1990s Labrador Sea convection, Nature Communications, 14, 4635, https://doi.org/10.1038/s41467-023-40323-9, 2023.

Caesar, L., McCarthy, G. D., Thornalley, D. J. R., Cahill, N., and Rahmstorf, S.: Current Atlantic Meridional Overturning Circulation weakest in last millennium, Nature Geoscience, 14, 118–120, https://doi.org/10.1038/s41561-021-00699-z, 2021.

Cunningham, S. A., Kanzow, T., Rayner, D., Baringer, M. O., Johns, W. E., Marotzke, J., Longworth, H. R., Grant, E. M., Hirschi, J. J.-M.,

900   Beal, L. M., Meinen, C. S., and Bryden, H. L.: Temporal Variability of the Atlantic Meridional Overturning Circulation at 26.5°N, Science, 317, 935–938, https://doi.org/10.1126/science.1141304, 2007.

Dale, M. and Fortin, M.-J.: From Graphs to Spatial Graphs, Annual Review of Ecology, Evolution, and Systematics, 41, 21–38, https://doi.org/10.1146/annurev-ecolsys-102209-144718, 2010.

Debreu, L., Vouland, C., and Blayo, E.: AGRIF: Adaptive grid refinement in Fortran, Computers Geosciences, 34, 8–13,

905   https://doi.org/https://doi.org/10.1016/j.cageo.2007.01.009, 2008.

Delman, A. and Landerer, F.: Downscaling Satellite-Based Estimates of Ocean Bottom Pressure for Tracking Deep Ocean Mass Transport, Remote Sensing, 14, 1764, https://doi.org/10.3390/rs14071764, 2022.

Dwivedi, R., Dave, D., Naik, H., Singhal, S., Omer, R., Patel, P., Qian, B., Wen, Z., Shah, T., Morgan, G., and Ranjan, R.: Explainable AI (XAI): Core Ideas, Techniques, and Solutions, ACM Computing Surveys, 55, 1–33, https://doi.org/10.1145/3561048, 2023.

910   Elipot, S., Frajka-Williams, E., Hughes, C. W., and Willis, J. K.: The Observed North Atlantic Meridional Overturning Circulation: Its Meridional Coherence and Ocean Bottom Pressure, Journal of Physical Oceanography, 44, 517–537, https://doi.org/10.1175/JPO-D-13-026.1, 2014.

Elipot, S., Frajka-Williams, E., Hughes, C. W., Olhede, S., and Lankhorst, M.: Observed Basin-Scale Response of the North Atlantic Meridional Overturning Circulation to Wind Stress Forcing, Journal of Climate, 30, 2029–2054, https://doi.org/10.1175/JCLI-D-16-0664.1,

915   2017.

Fisher, A., Rudin, C., and Dominici, F.: All models are wrong, but many are useful: Learning a variable's importance by studying an entire class of prediction models simultaneously, Journal of Machine Learning Research, 20, 1–81, 2019.

Forget, G., Campin, J.-M., Heimbach, P., Hill, C. N., Ponte, R. M., and Wunsch, C.: ECCO version 4: an integrated framework for non-linear inverse modeling and global ocean state estimation, Geoscientific Model Development, 8, 3071–3104, https://doi.org/10.5194/gmd-8-

920   3071-2015, 2015.

Fox, A. D., Handmann, P., Schmidt, C., Fraser, N., Rühs, S., Sanchez-Franks, A., Martin, T., Oltmanns, M., Johnson, C., Rath, W., Holliday, N. P., Biastoch, A., Cunningham, S. A., and Yashayaev, I.: Exceptional freshening and cooling in the eastern subpolar North Atlantic caused by reduced Labrador Sea surface heat loss, Ocean Science, 18, 1507–1533, https://doi.org/10.5194/os-18-1507-2022, 2022.

Frajka-Williams, E., Meinen, C. S., Johns, W. E., Smeed, D. A., Duchez, A., Lawrence, A. J., Cuthbertson, D. A., McCarthy, G. D., Bryden, H. L., Baringer, M. O., Moat, B. I., and Rayner, D.: Compensation between meridional flow components of the Atlantic MOC at 26° N, Ocean Science, 12, 481–493, https://doi.org/10.5194/os-12-481-2016, 2016.

Frajka-Williams, E., Ansorge, I. J., Baehr, J., Bryden, H. L., Chidichimo, M. P., Cunningham, S. A., Danabasoglu, G., Dong, S., Donohue, K. A., Elipot, S., Heimbach, P., Holliday, N. P., Hummels, R., Jackson, L. C., Karstensen, J., Lankhorst, M., Le Bras, I. A., Lozier, M. S., McDonagh, E. L., Meinen, C. S., Mercier, H., Moat, B. I., Perez, R. C., Piecuch, C. G., Rhein, M., Srokosz, M. A., Trenberth, K. E., Bacon, S., Forget, G., Goni, G., Kieke, D., Koelling, J., Lamont, T., McCarthy, G. D., Mertens, C., Send, U., Smeed, D. A., Speich, S., Van Den Berg, M., Volkov, D., and Wilson, C.: Atlantic Meridional Overturning Circulation: Observed Transport and Variability, Frontiers in Marine Science, 6, 260, https://doi.org/10.3389/fmars.2019.00260, 2019.

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

Hernández-Guerra, A., Joyce, T. M., Fraile-Nuez, E., and Vélez-Belchí, P.: Using Argo data to investigate the Meridional Overturning Circulation in the North Atlantic, Deep Sea Research Part I: Oceanographic Research Papers, 57, 29–36, https://doi.org/10.1016/j.dsr.2009.10.003, 2010.

Herrford, J., Brandt, P., Kanzow, T., Hummels, R., Araujo, M., and Durgadoo, J. V.: Seasonal variability of the Atlantic Meridional Overturning Circulation at 11° S inferred from bottom pressure measurements, Ocean Science, 17, 265–284, https://doi.org/10.5194/os-17-265-2021, 2021.

Heuzé, C.: Antarctic Bottom Water and North Atlantic Deep Water in CMIP6 models, Ocean Science, 17, 59–90, https://doi.org/10.5194/os-17-59-2021, 2021.

Hirschi, J. J., Barnier, B., Böning, C., Biastoch, A., Blaker, A. T., Coward, A., Danilov, S., Drijfhout, S., Getzlaff, K., Griffies, S. M., Hasumi, H., Hewitt, H., Iovino, D., Kawasaki, T., Kiss, A. E., Koldunov, N., Marzocchi, A., Mecking, J. V., Moat, B., Molines, J., Myers, P. G., Penduff, T., Roberts, M., Treguier, A., Sein, D. V., Sidorenko, D., Small, J., Spence, P., Thompson, L., Weijer, W., and Xu, X.: The Atlantic Meridional Overturning Circulation in High-Resolution Models, Journal of Geophysical Research: Oceans, 125, e2019JC015 522, https://doi.org/10.1029/2019JC015522, 2020.

Holte, J. and Straneo, F.: Seasonal Overturning of the Labrador Sea as Observed by Argo Floats, Journal of Physical Oceanography, 47, 2531–2543, https://doi.org/10.1175/JPO-D-17-0051.1, 2017.

Ioffe, S. and Szegedy, C.: Batch normalization: accelerating deep network training by reducing internal covariate shift, in: Proceedings of the 32nd International Conference on International Conference on Machine Learning - Volume 37, ICML'15, pp. 448–456, JMLR.org, place: Lille, France, 2015.

Jackson, L. C., Biastoch, A., Buckley, M. W., Desbruyères, D. G., Frajka-Williams, E., Moat, B., and Robson, J.: The evolution of the North Atlantic Meridional Overturning Circulation since 1980, Nature Reviews Earth & Environment, 3, 241–254, https://doi.org/10.1038/s43017-022-00263-2, 2022.

Jiang, W. and Luo, J.: Graph neural network for traffic forecasting: A survey, Expert Systems with Applications, 207, 117 921, https://doi.org/10.1016/j.eswa.2022.117921, 2022.

Johns, W. E., Beal, L. M., Baringer, M. O., Molina, J. R., Cunningham, S. A., Kanzow, T., and Rayner, D.: Variability of Shallow and Deep Western Boundary Currents off the Bahamas during 2004–05: Results from the 26°N RAPID–MOC Array, Journal of Physical Oceanography, 38, 605–623, https://doi.org/10.1175/2007JPO3791.1, 2008.

Johns, W. E., Baringer, M. O., Beal, L. M., Cunningham, S. A., Kanzow, T., Bryden, H. L., Hirschi, J. J. M., Marotzke, J., Meinen, C. S., Shaw, B., and Curry, R.: Continuous, Array-Based Estimates of Atlantic Ocean Heat Transport at 26.5°N, Journal of Climate, 24, 2429 – 2449, https://doi.org/10.1175/2010JCLI3997.1, 2011.

Johns, W. E., Elipot, S., Smeed, D. A., Moat, B., King, B., Volkov, D. L., and Smith, R. H.: Towards two decades of Atlantic Ocean mass and heat transports at 26.5° N, Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 381, 20220 188, https://doi.org/10.1098/rsta.2022.0188, 2023.

Johnson, G. C., Lyman, J. M., and Purkey, S. G.: Informing Deep Argo Array Design Using Argo and Full-Depth Hydrographic Section Data, Journal of Atmospheric and Oceanic Technology, 32, 2187–2198, https://doi.org/10.1175/JTECH-D-15-0139.1, 2015.

Kanzow, T., Hirschi, J. J. M., Meinen, C., Rayner, D., Cunningham, S. A., Marotzke, J., Johns, W. E., Bryden, H. L., Beal, L. M., and Baringer, M. O.: A prototype system for observing the Atlantic Meridional Overturning Circulation – scientific basis, measurement and risk mitigation strategies, and first results, Journal of Operational Oceanography, 1, 19–28, https://doi.org/10.1080/1755876X.2008.11020092, 2008.

Kanzow, T., Cunningham, S. A., Johns, W. E., Hirschi, J. J.-M., Marotzke, J., Baringer, M. O., Meinen, C. S., Chidichimo, M. P., Atkinson, C., Beal, L. M., Bryden, H. L., and Collins, J.: Seasonal Variability of the Atlantic Meridional Overturning Circulation at 26.5°N, Journal of Climate, 23, 5678–5698, https://doi.org/10.1175/2010JCLI3389.1, 2010.

Landerer, F. W., Flechtner, F. M., Save, H., Webb, F. H., Bandikova, T., Bertiger, W. I., Bettadpur, S. V., Byun, S. H., Dahle, C., Dobslaw, H., Fahnestock, E., Harvey, N., Kang, Z., Kruizinga, G. L. H., Loomis, B. D., McCullough, C., Murböck, M., Nagel, P., Paik, M., Pie, N., Poole, S., Strekalov, D., Tamisiea, M. E., Wang, F., Watkins, M. M., Wen, H., Wiese, D. N., and Yuan, D.: Extending the Global Mass Change Data Record: GRACE Follow-On Instrument and Science Data Performance, Geophysical Research Letters, 47, e2020GL088 306, https://doi.org/10.1029/2020GL088306, 2020.

Larsen, J.: Sixteen years of florida current transport at 27 N (2001) Geophys, Res. Lett, 28, 3179–3182, 2001.

Latif, M. and Keenlyside, N. S.: A perspective on decadal climate variability and predictability, Deep Sea Research Part II: Topical Studies in Oceanography, 58, 1880–1894, https://doi.org/10.1016/j.dsr2.2010.10.066, 2011.

Latif, M., Sun, J., Visbeck, M., and Hadi Bordbar, M.: Natural variability has dominated Atlantic Meridional Overturning Circulation since 1900, Nature Climate Change, 12, 455–460, https://doi.org/10.1038/s41558-022-01342-4, 2022.

Li, F., Lozier, M. S., Bacon, S., Bower, A. S., Cunningham, S. A., De Jong, M. F., deYoung, B., Fraser, N., Fried, N., Han, G., Holliday, N. P., Holte, J., Houpert, L., Inall, M. E., Johns, W. E., Jones, S., Johnson, C., Karstensen, J., Le Bras, I. A., Lherminier, P., Lin, X., Mercier, H., Oltmanns, M., Pacini, A., Petit, T., Pickart, R. S., Rayner, D., Straneo, F., Thierry, V., Visbeck, M., Yashayaev, I., and Zhou, C.: Subpolar North Atlantic western boundary density anomalies and the Meridional Overturning Circulation, Nature Communications, 12, 3002, https://doi.org/10.1038/s41467-021-23350-2, 2021.

Lindemann, B., Müller, T., Vietz, H., Jazdi, N., and Weyrich, M.: A survey on long short-term memory networks for time series prediction, Procedia CIRP, 99, 650–655, https://doi.org/10.1016/j.procir.2021.03.088, 2021.

Locarnini, R. A., Mishonov, A. V., Antonov, J. I., Boyer, T. P., Garcia, H. E., Baranova, O. K., Zweng, M. M., Paver, C. R., Reagan, J. R., Johnson, D. R., Hamilton, M., Seidov, 1948, D., and Levitus, S.: World ocean atlas 2013. Volume 1, Temperature, NOAA Atlas NESDIS series, https://doi.org/10.7289/V55X26VD, 2013.

Lozier, M. S.: Overturning in the North Atlantic, Annual Review of Marine Science, 4, 291–315, https://doi.org/10.1146/annurev-marine-120710-100740, 2012.

Markus, A. F., Kors, J. A., and Rijnbeek, P. R.: The role of explainability in creating trustworthy artificial intelligence for health care: A comprehensive survey of the terminology, design choices, and evaluation strategies, Journal of Biomedical Informatics, 113, 103 655, https://doi.org/10.1016/j.jbi.2020.103655, 2021.

McCarthy, G., Frajka-Williams, E., Johns, W. E., Baringer, M. O., Meinen, C. S., Bryden, H. L., Rayner, D., Duchez, A., Roberts, C., and Cunningham, S. A.: Observed interannual variability of the Atlantic meridional overturning circulation at 26.5°N, Geophysical Research Letters, 39, 2012GL052 933, https://doi.org/10.1029/2012GL052933, 2012.

McCarthy, G., Smeed, D., Johns, W., Frajka-Williams, E., Moat, B., Rayner, D., Baringer, M., Meinen, C., Collins, J., and Bryden, H.: Measuring the Atlantic Meridional Overturning Circulation at 26°N, Progress in Oceanography, 130, 91–111, https://doi.org/10.1016/j.pocean.2014.10.006, 2015.

McCarthy, G. D. and Caesar, L.: Can we trust projections of AMOC weakening based on climate models that cannot reproduce the past?, Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 381, 20220 193, https://doi.org/10.1098/rsta.2022.0193, 2023.

McCarthy, G. D., Hug, G., Smeed, D., Morris, K. J., and Moat, B.: Signal and Noise in the Atlantic Meridional Overturning Circulation at 26°N, Geophysical Research Letters, 52, e2025GL115 055, https://doi.org/10.1029/2025GL115055, 2025.

McMonigal, K., Beal, L. M., and Willis, J. K.: The Seasonal Cycle of the South Indian Ocean Subtropical Gyre Circulation as Revealed by Argo and Satellite Data, Geophysical Research Letters, 45, 9034–9041, https://doi.org/10.1029/2018GL078420, 2018.

Meilă, M. and Zhang, H.: Manifold Learning: What, How, and Why, Annual Review of Statistics and Its Application, 11, 393–417, https://doi.org/10.1146/annurev-statistics-040522-115238, 2024.

Moat, B. I., Smeed, D. A., Rayner, D., Johns, W. E., Smith, R., Volkov, D., Elipot, S., Petit, T., Kajtar, J., Baringer, M. O., and Collins, J.: Atlantic meridional overturning circulation observed by the RAPID-MOCHA-WBTS (RAPID-Meridional Overturning Circulation and Heatflux Array-Western Boundary Time Series) array at 26N from 2004 to 2023 (v2023.1), https://nora.nerc.ac.uk/id/eprint/538140/.

Moat, B. I., Smeed, D. A., Frajka-Williams, E., Desbruyères, D. G., Beaulieu, C., Johns, W. E., Rayner, D., Sanchez-Franks, A., Baringer, M. O., Volkov, D., Jackson, L. C., and Bryden, H. L.: Pending recovery in the strength of the meridional overturning circulation at 26° N, Ocean Science, 16, 863–874, https://doi.org/10.5194/os-16-863-2020, 2020.

Petit, T., Lozier, M. S., Rühs, S., Handmann, P., and Biastoch, A.: Propagation and Transformation of Upper North Atlantic Deep Water From the Subpolar Gyre to 26.5°N, Journal of Geophysical Research: Oceans, 128, e2023JC019 726, https://doi.org/10.1029/2023JC019726, 2023.

Reeve, K. A., Boebel, O., Strass, V., Kanzow, T., and Gerdes, R.: Horizontal circulation and volume transports in the Weddell Gyre derived from Argo float data, Progress in Oceanography, 175, 263–283, https://doi.org/10.1016/j.pocean.2019.04.006, 2019.

Riser, S. C., Freeland, H. J., Roemmich, D., Wijffels, S., Troisi, A., Belbéoch, M., Gilbert, D., Xu, J., Pouliquen, S., Thresher, A., Le Traon, P.-Y., Maze, G., Klein, B., Ravichandran, M., Grant, F., Poulain, P.-M., Suga, T., Lim, B., Sterl, A., Sutton, P., Mork, K.-A., Vélez-Belchí, P. J., Ansorge, I., King, B., Turton, J., Baringer, M., and Jayne, S. R.: Fifteen years of ocean observations with the global Argo array, Nature Climate Change, 6, 145–153, https://doi.org/10.1038/nclimate2872, 2016.

Roemmich, D., Johnson, G. C., Riser, S., Davis, R., Gilson, J., Owens, W. B., Garzoli, S. L., Schmid, C., and Ignaszweski, M.: The Argo Program: Observing the Global Ocean with Profiling Floats, Oceanography, 22, 34–43, http://www.jstor.org/stable/24860957, publisher: Oceanography Society, 2009.

Roemmich, D., Alford, M. H., Claustre, H., Johnson, K., King, B., Moum, J., Oke, P., Owens, W. B., Pouliquen, S., Purkey, S., Scanderbeg, M., Suga, T., Wijffels, S., Zilberman, N., Bakker, D., Baringer, M., Belbeoch, M., Bittig, H. C., Boss, E., Calil, P., Carse, F., Carval, T., Chai, F., Conchubhair, D. , d'Ortenzio, F., Dall'Olmo, G., Desbruyeres, D., Fennel, K., Fer, I., Ferrari, R., Forget, G., Freeland, H., Fujiki, T., Gehlen, M., Greenan, B., Hallberg, R., Hibiya, T., Hosoda, S., Jayne, S., Jochum, M., Johnson, G. C., Kang, K., Kolodziejczyk, N., Körtzinger, A., Traon, P.-Y. L., Lenn, Y.-D., Maze, G., Mork, K. A., Morris, T., Nagai, T., Nash, J., Garabato, A. N., Olsen, A., Pattabhi, R. R., Prakash, S., Riser, S., Schmechtig, C., Schmid, C., Shroyer, E., Sterl, A., Sutton, P., Talley, L., Tanhua, T., Thierry, V., Thomalla, S., Toole, J., Troisi, A., Trull, T. W., Turton, J., Velez-Belchi, P. J., Walczowski, W., Wang, H., Wanninkhof, R., Waterhouse, A. F., Waterman, S., Watson, A., Wilson, C., Wong, A. P. S., Xu, J., and Yasuda, I.: On the Future of Argo: A Global, Full-Depth, Multi-Disciplinary Array, Frontiers in Marine Science, 6, https://doi.org/10.3389/fmars.2019.00439, 2019.

Rühs, S., Oliver, E. C. J., Biastoch, A., Böning, C. W., Dowd, M., Getzlaff, K., Martin, T., and Myers, P. G.: Changing Spatial Patterns of Deep Convection in the Subpolar North Atlantic, Journal of Geophysical Research: Oceans, 126, e2021JC017 245, https://doi.org/10.1029/2021JC017245, 2021.

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

Schiller-Weiss, I., Schulzki, T., Schwarzkopf, F. U., Martin, T., and Biastoch, A.: Sea Surface Height Response to Decadal-Scale AMOC Changes in an Eddy-Rich Ocean Model, Journal of Climate, 38, 5265–5283, https://doi.org/10.1175/JCLI-D-24-0646.1, 2025.

Sinha, B., Smeed, D., McCarthy, G., Moat, B., Josey, S., Hirschi, J.-M., Frajka-Williams, E., Blaker, A., Rayner, D., and Madec, G.: The accuracy of estimates of the overturning circulation from basin-wide mooring arrays, Progress in Oceanography, 160, 101–123, https://doi.org/10.1016/j.pocean.2017.12.001, 2018.

Smeed, D. A., Josey, S. A., Beaulieu, C., Johns, W. E., Moat, B. I., Frajka-Williams, E., Rayner, D., Meinen, C. S., Baringer, M. O., Bryden, H. L., and McCarthy, G. D.: The North Atlantic Ocean Is in a State of Reduced Overturning, Geophysical Research Letters, 45, 1527–1533, https://doi.org/10.1002/2017GL076350, 2018.

Solodoch, A., Stewart, A. L., McC. Hogg, A., and Manucharyan, G. E.: 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, e2022MS003 370, https://doi.org/10.1029/2022MS003370, 2023.

Sonnewald, M. and Lguensat, R.: Revealing the Impact of Global Heating on North Atlantic Circulation Using Transparent Machine Learning, Journal of Advances in Modeling Earth Systems, 13, e2021MS002 496, https://doi.org/10.1029/2021MS002496, 2021.

Terhaar, J., Vogt, L., and Foukal, N. P.: Atlantic overturning inferred from air-sea heat fluxes indicates no decline since the 1960s, Nature Communications, 16, 222, https://doi.org/10.1038/s41467-024-55297-5, 2025.

Tsujino, H., Urakawa, S., Nakano, H., Small, R. J., Kim, W. M., Yeager, S. G., Danabasoglu, G., Suzuki, T., Bamber, J. L., Bentsen, M., Böning, C. W., Bozec, A., Chassignet, E. P., Curchitser, E., Boeira Dias, F., Durack, P. J., Griffies, S. M., Harada, Y., Ilicak, M., Josey, S. A., Kobayashi, C., Kobayashi, S., Komuro, Y., Large, W. G., Le Sommer, J., Marsland, S. J., Masina, S., Scheinert, M., Tomita, H., Valdivieso, M., and Yamazaki, D.: JRA-55 based surface dataset for driving ocean–sea-ice models (JRA55-do), Ocean Modelling, 130, 79–139, https://doi.org/https://doi.org/10.1016/j.ocemod.2018.07.002, 2018.

Tsujino, H., Urakawa, L. S., Griffies, S. M., Danabasoglu, G., Adcroft, A. J., Amaral, A. E., Arsouze, T., Bentsen, M., Bernardello, R., Böning, C. W., Bozec, A., Chassignet, E. P., Danilov, S., Dussin, R., Exarchou, E., Fogli, P. G., Fox-Kemper, B., Guo, C., Ilicak, M., Iovino, D., Kim, W. M., Koldunov, N., Lapin, V., Li, Y., Lin, P., Lindsay, K., Liu, H., Long, M. C., Komuro, Y., Marsland, S. J., Masina, S., Nummelin, A., Rieck, J. K., Ruprich-Robert, Y., Scheinert, M., Sicardi, V., Sidorenko, D., Suzuki, T., Tatebe, H., Wang, Q., Yeager, S. G., and

Yu, Z.: Evaluation of global ocean–sea-ice model simulations based on the experimental protocols of the Ocean Model Intercomparison Project phase 2 (OMIP-2), Geoscientific Model Development, 13, 3643–3708, https://doi.org/10.5194/gmd-13-3643-2020, 2020.

van der Maaten, L. and Hinton, G.: Visualizing Data using t-SNE, Journal of Machine Learning Research, 9, 2579–2605, http://jmlr.org/papers/v9/vandermaaten08a.html, 2008.

Wen, Q., Zhou, T., Zhang, C., Chen, W., Ma, Z., Yan, J., and Sun, L.: Transformers in time series: A survey, in: International Joint Conference on Artificial Intelligence(IJCAI), 2023.

Willis, J. K.: Can in situ floats and satellite altimeters detect long-term changes in Atlantic Ocean overturning?, Geophysical Research Letters, 37, 2010GL042 372, https://doi.org/10.1029/2010GL042372, 2010.

Wölker, Y., Beth, C., Renz, M., and Biastoch, A.: SUSTeR: Sparse Unstructured Spatio Temporal Reconstruction on Traffic Prediction, in: Proceedings of the 31st ACM International Conference on Advances in Geographic Information Systems, SIGSPATIAL '23, Association for Computing Machinery, New York, NY, USA, ISBN 9798400701689, https://doi.org/10.1145/3589132.3625631, 2023.

Wong, A. P. S., Wijffels, S. E., Riser, S. C., Pouliquen, S., Hosoda, S., Roemmich, D., Gilson, J., Johnson, G. C., Martini, K., Murphy, D. J., Scanderbeg, M., Bhaskar, T. V. S. U., Buck, J. J. H., Merceur, F., Carval, T., Maze, G., Cabanes, C., André, X., Poffa, N., Yashayaev, I., Barker, P. M., Guinehut, S., Belbéoch, M., Ignaszewski, M., Baringer, M. O., Schmid, C., Lyman, J. M., McTaggart, K. E., Purkey, S. G., Zilberman, N., Alkire, M. B., Swift, D., Owens, W. B., Jayne, S. R., Hersh, C., Robbins, P., West-Mack, D., Bahr, F., Yoshida, S., Sutton, P. J. H., Cancouët, R., Coatanoan, C., Dobbler, D., Juan, A. G., Gourrion, J., Kolodziejczyk, N., Bernard, V., Bourlès, B., Claustre, H., D'Ortenzio, F., Le Reste, S., Le Traon, P.-Y., Rannou, J.-P., Saout-Grit, C., Speich, S., Thierry, V., Verbrugge, N., Angel-Benavides, I. M., Klein, B., Notarstefano, G., Poulain, P.-M., Vélez-Belchí, P., Suga, T., Ando, K., Iwasaska, N., Kobayashi, T., Masuda, S., Oka, E., Sato, K., Nakamura, T., Sato, K., Takatsuki, Y., Yoshida, T., Cowley, R., Lovell, J. L., Oke, P. R., Van Wijk, E. M., Carse, F., Donnelly, M., Gould, W. J., Gowers, K., King, B. A., Loch, S. G., Mowat, M., Turton, J., Rama Rao, E. P., Ravichandran, M., Freeland, H. J., Gaboury, I., Gilbert, D., Greenan, B. J. W., Ouellet, M., Ross, T., Tran, A., Dong, M., Liu, Z., Xu, J., Kang, K., Jo, H., Kim, S.-D., and Park, H.-M.: Argo Data 1999–2019: Two Million Temperature-Salinity Profiles and Subsurface Velocity Observations From a Global Array of Profiling Floats, Frontiers in Marine Science, 7, 700, https://doi.org/10.3389/fmars.2020.00700, 2020.

Wu, Q., Jochum, M., Avery, J. E., Vettoretti, G., and Nuterman, R.: Machine Guided Derivation of the Atlantic Meridional Overturning Circulation (AMOC) Strength, Geophysical Research Letters, 52, e2024GL113 454, https://doi.org/10.1029/2024GL113454, 2025.

Wölker, Y.: ywoelker/amoc_reconstruction: AMOC Reconstruction v.0.1.0 [Submission of Manuscript], https://doi.org/10.5281/zenodo.15643942, 2025a.

Wölker, Y.: ywoelker/AMOC-Reconstruction-from-Argo-profiles: Estimating AMOC from Argo Profiles with Machine Learning Trained on Ocean Simulations [Submission], https://doi.org/10.5281/zenodo.15644138, 2025b.

Wölker, Y.: Estimating AMOC from Argo Profiles with Machine Learning Trained on Ocean Simulations [Experiment Results], https://doi.org/10.5281/zenodo.15644013, 2025c.

Wölker, Y., Rath, W., and Biastoch, A.: Estimating AMOC from Argo Profiles with Machine Learning Trained on Ocean Simulations [Dataset], https://doi.org/10.5281/zenodo.15643439, 2025.

Yan, H., Ma, X., and Pu, Z.: Learning Dynamic and Hierarchical Traffic Spatiotemporal Features With Transformer, IEEE Transactions on Intelligent Transportation Systems, 23, 22 386–22 399, https://doi.org/10.1109/TITS.2021.3102983, 2022.

Zhang, R., Sutton, R., Danabasoglu, G., Kwon, Y., Marsh, R., Yeager, S. G., Amrhein, D. E., and Little, C. M.: A Review of the Role of the Atlantic Meridional Overturning Circulation in Atlantic Multidecadal Variability and Associated Climate Impacts, Reviews of Geophysics, 57, 316–375, https://doi.org/10.1029/2019RG000644, 2019.

1110   Zilberman, N. V., Roemmich, D. H., and Gilson, J.: Deep-Ocean Circulation in the Southwest Pacific Ocean Interior: Estimates of the Mean Flow and Variability Using Deep Argo Data, Geophysical Research Letters, 47, e2020GL088 342, https://doi.org/10.1029/2020GL088342, 2020.

Zilberman, N. V., Scanderbeg, M., Gray, A. R., and Oke, P. R.: Scripps Argo Trajectory-Based Velocity Product: Global Estimates of Absolute Velocity Derived from Core, Biogeochemical, and Deep Argo Float Trajectories at Parking Depth, Journal of Atmospheric and Oceanic

1115   Technology, 40, 361–374, https://doi.org/10.1175/JTECH-D-22-0065.1, 2023.

Zweng, M. M., Reagan, J. R., Antonov, J. I., Locarnini, R. A., Mishonov, A. V., Boyer, T. P., Garcia, H. E., Baranova, O. K., Johnson, D. R., Seidov, 1948, D., Biddle, M. M., and Levitus, S.: World ocean atlas 2013. Volume 2, Salinity, NOAA Atlas NESDIS series, https://doi.org/10.7289/V5251G4D, 2013.