JGR | Machine Learning and Computation

RESEARCH ARTICLE
10.1029/2025JH001182

Basin-Wide Atlantic Ocean Water Mass Classification and Climatic Variability From Machine Learning

Joshua Lanham[1], **Kaushik Srinivasan**[2], **Laura Cimoli**[3], **and Ali Mashayek**[1,3,4]

[1]Department of Earth Sciences, University of Cambridge, Cambridge, UK, [2]Atmospheric and Oceanic Sciences, University of California Los Angeles, Los Angeles, CA, USA, [3]Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Cambridge, UK, [4]National Oceanography Centre, Southampton, UK

Key Points:

- A random forest ensemble extends ocean water-mass classification from limited tracer observations to abundant temperature-salinity data
- The model classifies Atlantic water masses with high skill (R^2 = 0.8–0.95) in spatially independent tests across most of the basin
- Applying the model to an ocean state estimate enables 30-year trend estimates of water masses shaping the Atlantic overturning circulation

Supporting Information:

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

Correspondence to:

J. Lanham,
jal238@cam.ac.uk

Citation:

Lanham, J., Srinivasan, K., Cimoli, L., & Mashayek, A. (2026). Basin-wide Atlantic Ocean water mass classification and climatic variability from machine learning. *Journal of Geophysical Research: Machine Learning and Computation*, *3*, e2025JH001182. https://doi.org/10.1029/2025JH001182

Received 10 DEC 2025
Accepted 3 APR 2026

Abstract Identification of water masses in the Atlantic Ocean is key to understanding large-scale circulation, transport, and mixing processes. However, traditional classification methods, such as Optimum Multi-Parameter analysis (OMP), are often limited by relatively sparse hydrographic profiles. Here, we develop a hybrid framework, which uses a random forest (RF) modeling approach trained upon an initial OMP analysis that is itself fully constrained by a range of biogeochemical tracers. The resulting model performs robustly even in the absence of such tracers. Given that several observational platforms measure temperature and salinity only, this approach enables the skillful classification of water masses within a much larger expanse of observational data. It also facilitates water mass analysis within large-scale state-estimate products and model output. We apply our RF model ensemble to the Estimating the Circulation and Climate of the Ocean (ECCO) state estimate to produce a gridded Atlantic Ocean water mass product at monthly resolution, which we use to infer changes in Atlantic water mass structure over recent decades. Results indicate a contraction in Antarctic Bottom Water, an expansion of Central Water at the expense of Antarctic Intermediate Water in the Southern Ocean, and a possible poleward shift in Circumpolar Deep Water.

Plain Language Summary Water masses are large bodies of water with distinct properties. Identifying them helps us understand how the ocean moves, mixes, and transports heat, carbon, oxygen, and other properties. This usually requires detailed chemical measurements, which are typically only available in few sparse locations along ship tracks. In this study, we train a machine learning model using those detailed measurements in the Atlantic Ocean. Once trained, the model can identify water masses using just temperature and salinity data that are much more widely available, enabling the mapping of water masses across the Atlantic Ocean at much higher spatial and temporal resolution. We apply this machine learning model to a data-assimilating ocean model to produce an estimate of the water mass distribution in the Atlantic Ocean every month between 1992 and 2018 and use it to infer changes in the water mass structure, theoretically enabling insights into AMOC variability and tipping points.

1. Introduction

Water masses provide a practical framework for describing the ocean's interior structure and for quantifying the transport and storage of heat, carbon, nutrients, and freshwater. They are characterized by distinct, quasi-conserved property signatures set at formation and subsequently modified primarily through mixing and non-conservative biogeochemical processes along circulation pathways (Talley, 2011).

Robust classification of water masses is necessary to understand the large-scale circulation, its role in global climate regulation via the transport of heat and other properties, and its response to anthropogenic climate change. Because mixing between water masses governs the transformation pathways that set the structure of the overturning circulation, classification frameworks must capture how distinct water masses are modified along their flow paths (Cimoli, Mashayek, et al., 2023; Mashayek et al., 2015). In addition, water masses act as carriers of heat, carbon, nutrients, and freshwater, so identifying them consistently is fundamental for tracer budgets across a range of spatial and temporal scales (Cimoli, Gebbie, et al., 2023; Ellison et al., 2023).

The established framework for water mass identification is Optimum Multi-Parameter (OMP) analysis. OMP identifies water masses as linear mixtures of predefined source water types (SWTs), characterized by distinct thermohaline and biogeochemical property signatures at formation (Thompson & Edwards, 1981; Tomczak, 1981). In practice, OMP classifications are typically constrained using biogeochemical tracers such as

oxygen and nutrients, which provide essential discriminatory power between water masses with similar thermohaline properties. As a result, OMP analysis is largely limited to tracer-rich hydrographic data sets, including ship-based bottle measurements from programmes such as Global Ocean Ship-based Hydrographic Investigations Program (GO-SHIP) (Sloyan et al., 2019) and WOCE, or gridded climatologies such as the Global Ocean Data Analysis Project (GLODAP) (Lauvset et al., 2016).

However, there is a much larger fraction of ocean observational platforms and modeling products that return measurements of temperature and salinity only, that is, in which measurements of such tracers are absent. This includes the vast majority of Argo floats, ice-tethered profilers, CTD-only ship casts, and all ocean models without a biogeochemical component (see data coverage in Text S1, Figure S1 of Supporting Information S1). This means that basin-scale and time-continuous water mass fields are difficult to obtain, motivating the development of new approaches that can extend classification to the far larger volume of temperature-salinity (T-S) only observations and model output. Machine learning techniques offer an opportunity to expand the range of classifiable data and thereby improve the spatio-temporal resolution of water mass products.

Water mass identification and tracking are central to diagnosing the structure and variability of the large-scale overturning circulation. In the Atlantic, the formation of North Atlantic Deep Water (NADW) in the Labrador and Greenland-Iceland-Norway seas ventilates the upper branch of the Atlantic Meridional Overturning Circulation (AMOC), exporting heat and salinity anomalies southward at mid-depths. In the Southern Ocean, dense shelf water production and deep convection in the Weddell Sea form Antarctic Bottom Water (AABW), which supplies the abyssal limb of the AMOC and constitutes the primary ventilation pathway for the deep global ocean. In the Southern Ocean, the mixing of NADW at depth with AABW, Pacific Deep Water, and Indian Ocean Deep Water forms Circumpolar Deep Water (CDW), which upwells along tilted isopycnals to the Antarctic continent. Intrusions of the relatively warm CDW onto the Antarctic continental shelf are thought to be the principal agent of the basal melting of Antarctic ice shelves (Jacobs et al., 1992; Pritchard et al., 2012). In recent decades, observations indicate that AABW has warmed, freshened, and contracted (Anilkumar et al., 2021; Gunn et al., 2023; Johnson & Purkey, 2024; Kobayashi, 2018; Purkey & Johnson, 2010, 2012; Silvano et al., 2023), while model projections suggest that enhanced buoyancy forcing in subpolar North Atlantic formation regions may substantially weaken or even collapse the upper cell of the AMOC (Caesar et al., 2021; Ditlevsen & Ditlevsen, 2023; Rahmstorf et al., 2015). These changes directly affect the abyssal storage of heat and carbon, the rate of deep-ocean ventilation, and the stability of the global climate system.

In addition to these deep waters, intermediate and upper layers provide key pathways for the redistribution of heat, freshwater, and biogeochemical tracers. In the Southern Ocean, wind-driven subduction north of the Antarctic Circumpolar Current (ACC) forms Antarctic Intermediate Water (AAIW), which spreads equatorward at depths of 500–1,500 m and ventilates the lower thermocline. In the North Atlantic, Subarctic Intermediate Water (SAIW) forms on the western flank of the subpolar gyre and contributes to intergyre exchange (Liu & Tanhua, 2021). At shallower depths, central waters (CW) and mode waters are generated through wintertime mixed-layer deepening and subduction, supplying the thermocline with recently ventilated water. These intermediate and central/mode waters act as reservoirs for excess heat and anthropogenic carbon, while also regulating nutrient supply to the low-latitude thermocline, thereby influencing primary productivity (Gao et al., 2018; Z. Li et al., 2023; Sloyan & Rintoul, 2001; Sarmiento et al., 2004). The vertical and meridional structure of these water masses is schematically illustrated on the Atlantic overturning stream function in Figure 1.

Recent work has demonstrated that supervised machine-learning algorithms can emulate OMP-derived water mass fractions. Romera-Castillo et al. (2022) applied an ensemble regression method to a limited number of Atlantic hydrographic sections, showing that OMP and machine learning could be combined to classify water masses from ship section data in the absence of biogeochemical tracers. Here, we extend this approach by training a random forest (RF) model ensemble on the full set of tracer-rich GLODAP Atlantic bottle profiles, enabling basin-wide, gridded water mass classification rather than section-based estimates. We further demonstrate that this framework can be applied directly to a data-assimilating ocean state estimate (ECCO), yielding a continuous three-dimensional water mass product at monthly resolution across the Atlantic basin.

We run the analysis in three stages, as shown schematically in Figure 1. First, we run an OMP-based water mass classification in the Atlantic Ocean by using tracer-rich GLODAP cruise profiles. Second, we train a RF model ensemble on the OMP output, but using only conservative temperature, absolute salinity, and basic positional information (latitude, longitude, height above the bottom, and depth). Last, we apply the RF model to the ECCO

10.1029/2025JH001182

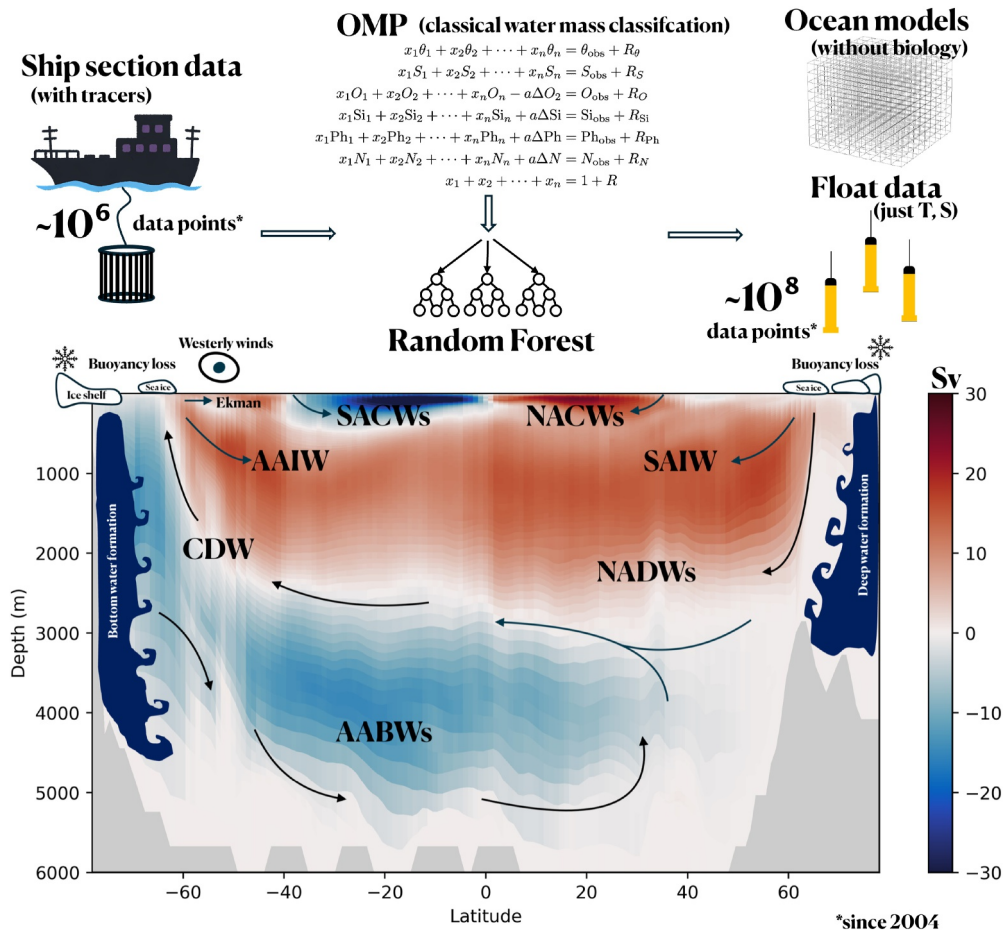

Figure 1. Extending classical OMP water mass classification with machine learning, and the large-scale water mass structure of the Atlantic Ocean. The top panels demonstrate schematically the method used in this study. CTD and biogeochemical bottle observations along ship transects (top left) are collated and used to identify water masses via a least-squares linear mixing solution known as OMP analysis. We train a RF ensemble on this classification (top middle), which enables application to a much wider range of thermohaline-only data sets (top right, see Text S1, Figure S1 in Supporting Information S1). Arrows indicate the sequential workflow. The bottom panel plots the global zonally integrated overturning stream function, calculated from ECCO (ECCO Consortium et al., 2021; Forget et al., 2015). Red contours indicate clockwise transport and blue indicates anti-clockwise transport. Units are given in Sverdrups, where 1 Sv is equivalent to $10^6 m^3 s^{-1}$. The water masses that are classified in this study are overlaid.

state estimate to produce a fully gridded monthly water mass product for the period 1992–2018. The methodology is described in detail below (Section 2), followed by the Results per each step (Section 3) and a final Discussion (Section 4).

2. Data and Methods

2.1. Initial OMP Classification

We begin by conducting an OMP analysis to classify the large-scale water masses in the Atlantic Ocean. OMP analysis assumes that each water parcel is a linear mixture of predefined end members, with the relative contributions constrained to be non-negative and (approximately) to sum to one. The analysis is based on solving a system of linear mixing equations using a least-squares optimization, where the residual between the observed property vector and the reconstructed mixture is minimized (Tomczak, 1981). Additionally, we apply a set of user-defined weights (shown in Table 1) to each variable to reflect differences in measurement uncertainty (Lauvset et al. (2024) reports GLODAPv2 basin-scale inter-cruise consistency of approximately 0.005 in salinity, 1% in oxygen, and 2% in nutrients). As in most OMP applications, measurement uncertainties in the underlying

Table 1
Properties of SWT End Members Used in This Study

SWT	Con. Temp.(°C) [24]	Abs. Sal.(g/kg) [24]	O_2 (μmol/kg) [7]	NO_3 (μmol/kg) [2]	PO_4 (μmol/kg) [2]	$Si(OH)_4$ (μmol/kg) [2]
CW	14.13	35.71	216.2	8.0	0.56	3.2
AAIW	1.78	33.82	300.7	27.3	1.95	21.1
SAIW	3.62	34.99	294.6	15.6	1.04	8.5
uNADW	3.33	35.07	279.4	17.0	1.11	11.4
lNADW	2.96	35.08	278.0	16.8	1.10	13.2
CDW	0.41	34.85	203.8	33.5	2.31	115.5
AABW	−0.46	34.83	239.0	32.8	2.27	124.9

Note. The water masses are Central Water (CW), Antarctic Intermediate Water (AAIW), Sub-Arctic Intermediate Water (SAIW), upper North Atlantic Deep Water (uNADW), lower North Atlantic Deep Water (lNADW), Circumpolar Deep Water (CDW), and Antarctic Bottom Water (AABW). Parameter weights (shown in square brackets) specify the relative importance assigned to each tracer in the OMP least-squares mixing formulation, with higher weights enforcing closer agreement between observed and reconstructed values.

hydrographic and biogeochemical tracers are therefore handled implicitly through the weighted least-squares formulation, rather than through explicit uncertainty propagation. The end-member values listed in Table 1 should thus be interpreted as representative mean properties. Further details, including all relevant equations, are provided in the Text S2 of Supporting Information S1.

In addition to conservative tracers such as conservative temperature and absolute salinity, the analysis includes quasi-conservative biogeochemical variables that provide discriminatory power between water masses. Temperature and salinity typically contain most of the information required for water mass identification, but additional tracers are valuable when water masses exhibit similar thermohaline properties yet differ in their biogeochemical signatures (Johnson, 2008). Rather than explicitly forming the classical composite tracers PO and NO, we follow a semi-conservative approach in which oxygen, phosphate, and nitrate are coupled through fixed remineralization exchange ratios derived from Redfield stoichiometry (Broecker, 1974; Karstensen & Tomczak, 1998). These exchange ratios provide the same discriminatory separation between water masses with similar thermohaline characteristics but distinct biogeochemical signatures, while allowing for non-conservative behavior to be represented through so-called ΔQ terms. Using the `pyompa` implementation from Shrikumar et al. (2022), we impose the physical constraints of non-negativity and a hard mass constraint (see Text in Supporting Information S1for further details).

The observational data for this classification come from the ship-based GLODAPv2 compiled data set. GLODAP provides a quality-controlled, internally consistent, and bias-adjusted compilation of discrete bottle data from ship-based hydrographic cruises across the global ocean (Olsen et al., 2020). Importantly, it includes the full suite of biogeochemical variables necessary to constrain the water mass classification problem. All GLODAP data used in this study are shown in Figure 2a.

For the purposes of this study, we adopt an idealized set of seven basin-scale source water types to represent the large-scale water mass structure of the Atlantic Ocean, chosen to ensure that the basin-wide OMP system remains well constrained. Source water properties (end members) are taken from Liu and Tanhua (2021). Where they define multiple sub-types of a given water mass (e.g., multiple central waters), we adopt a single representative end member by taking the mean of the corresponding source water properties. The resulting set of seven end members is summarized in Table 1.

To ensure basin-wide consistency and avoid the need to partition the Atlantic Ocean into separate regions, we apply this fixed set of seven end members uniformly across the basin. This choice differs from the approach of Liu and Tanhua (2021), who subdivide the Atlantic into multiple regions and solve locally restricted OMP systems in order to represent a larger number of water mass sub-types. Similarly, we note that the original Romera-Castillo et al. (2019) OMP framework used by Romera-Castillo et al. (2022) also applies end-member mixing over a domain that is divided into separate regions. While that approach is well suited to section-based analyses, it introduces discontinuities in water mass fractions at sub-domain boundaries that are problematic for gridded, basin-scale applications.

10.1029/2025JH001182

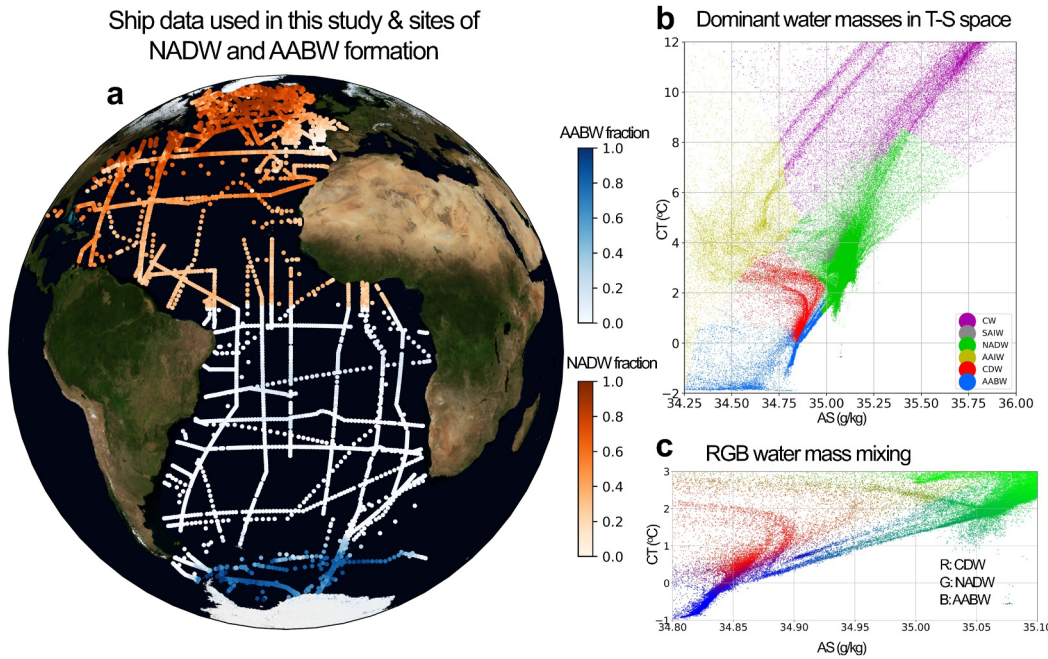

Figure 2. Spatial distribution and thermohaline characteristics of the GLODAP data set used in this study. Panel (a) shows the locations of GLODAP stations used in this study, overlain on an orthographic projection of the Atlantic Ocean. Shading indicates the fraction of NADW at 1,000 m in the Northern Hemisphere and the fraction of AABW at 2,000 m in the Southern Hemisphere. Panel (b) shows conservative temperature–absolute salinity space colored by the dominant water mass at each point, defined as the maximum OMP-derived fraction (hard classification). Panel (c) shows conservative temperature–absolute salinity space colored using an RGB blending approach, with CDW mapped to red, NADW to green, and AABW to blue. In panels (b, c), the upper and lower components of NADW are grouped.

Because our random-forest model ensemble is trained directly on OMP-derived water mass fractions, any discontinuities in the training data would be inherited by the machine-learning predictions. Adopting a single, basin-wide set of end members therefore ensures spatially coherent water mass fields that are suitable for extension to gridded data sets and ocean state estimates. Although this simplification necessarily limits the number of distinct water masses represented in order to keep the mixing problem formally determined, underdetermined OMP formulations are possible and widely used when additional constraints or regularization are imposed. Here we retain the simpler, well-constrained configuration to prioritize robustness and methodological clarity at the basin scale. The OMP solution of Liu and Tanhua (2021) is therefore not used directly in our primary analysis, but is included in the Supporting Information S1 as a qualitative point of comparison (Figures S2–S4 in Supporting Information S1).

2.2. Random Forest Model

We use a RF model to predict water mass fractions from hydrographic properties. A RF is a supervised machine-learning method that combines the predictions of many decision trees, where each tree learns a set of simple, rule-based splits in the input data. This makes RF models well suited to tabular geophysical data sets, such as hydrographic observations, where relationships between variables are non-linear and water masses are separated by relatively sharp boundaries in property space.

We choose a RF as the predictive model because it provides a straightforward way to predict multiple water mass fractions simultaneously within a single model (multi-output regression). In contrast, some alternative tree-based methods such as gradient boosting typically require training a separate model for each target variable, increasing complexity. RFs also require relatively little hyperparameter tuning, meaning that model performance is less sensitive to the precise choice of user-defined settings (such as the number or depth of trees).

Because RFs are based on ensembles of decision trees, they are particularly effective at capturing non-linear relationships and sharp transitions in feature space, which are characteristic of boundaries between different water masses. Previous benchmarking studies have shown that tree-based models such as RFs often outperform neural networks for tabular data sets of this kind, where inputs are structured physical variables rather than images or sequences (Grinsztajn et al., 2022). An explicit demonstration of this behavior in an oceanographic context is provided by Mashayek et al. (2022).

To train the RF model, we use a stratified 5-fold cross-validation approach, outlined as follows:

1. The OMP output training set is prepared, shuffled and split evenly into five partitions (folds). Each fold is used once as a validation set, while the remaining four serve as the training data.
2. For each fold, we train a distinct RF regressor on its training subset. Each RF consists of an ensemble of decision trees with a maximum depth of 16, and the number of trees is selected through hyperparameter tuning. The trained model is then evaluated on its respective test fold.
3. We quantify model performance using the coefficient of determination, R^2, which measures the fraction of variance in the OMP-derived water mass fractions explained by the model predictions. We calculate the R^2 score for each fold's test set, providing a reliable estimate of model accuracy through cross-validation covering the whole data set.
4. This process yields five separate RF ensemble models. Each of these may now be applied to the required data set. At each grid cell or observed data point, the final prediction is the average of the five model outputs, and the spread between them provides an estimate of uncertainty.

Initially, we undertake this process using temperature, salinity, positional data, and all available biogeochemical tracers as model features. To assess how well the model generalizes beyond the training data, we evaluate its performance using an out-of-distribution testing approach. The Atlantic Ocean is divided into a set of geographical boxes, and all GLODAP profiles within each box are sequentially withheld from the training set. For each iteration, the RF model is trained on data from all remaining boxes and subsequently tested on the withheld box. This ensures that the model is validated against spatially independent data (as opposed to random sub-samples), avoiding data leakage from nearby or overlapping profiles and providing a more stringent test of the model's ability to generalize across regions. Quantitative evaluation of model performance using this approach is presented in Section 3.2.

The trained RF ensemble is subsequently applied to gridded ocean state estimates in order to generate basin-wide, time-continuous water mass fields. For this purpose, we use the ECCOv4r4 ocean state estimate, which combines a general circulation model with a global data assimilation framework. ECCO assimilates a wide range of observational constraints, including satellite altimetry, Argo temperature and salinity profiles, GRACE gravity data, and hydrographic observations, to produce a dynamically consistent estimate of the ocean state (ECCO Consortium et al., 2021; Forget et al., 2015).

We choose ECCO rather than an interpolated observational product or an unconstrained reanalysis because it provides a physically consistent, mass and energy-conserving three-dimensional field, allowing water mass pathways and transformations to be interpreted coherently. Importantly for this study, ECCO does not include biogeochemical tracers, making it an appropriate testbed for evaluating the ability of the RF model to extend tracer-informed water mass classifications to temperature and salinity-only data sets.

Integrated water mass layer thicknesses are computed within ECCO at each timestep as: $H = \sum_k f_k \, \Delta z_k$, where f_k is the water mass fraction at vertical level k and Δz_k is the corresponding layer thickness. Changes in layer thickness are quantified as the difference between the mean of the final 5 years and the mean of the first 5 years of the ECCO period (1992–2018).

3. Results

3.1. OMP-Based Reference Classification of Atlantic Water Masses

Figure 3 shows meridional cross sections of Atlantic water mass fractions derived from the OMP classification. AABW is concentrated near the Antarctic margin in the Weddell Sea sector and is primarily confined to abyssal depths below 4,000 m (Figure 3a). NADW occupies a broad mid-depth layer between approximately 1,000 and 4,000 m, extending from the subpolar North Atlantic to around 30°S (Figure 3b), consistent with its formation in

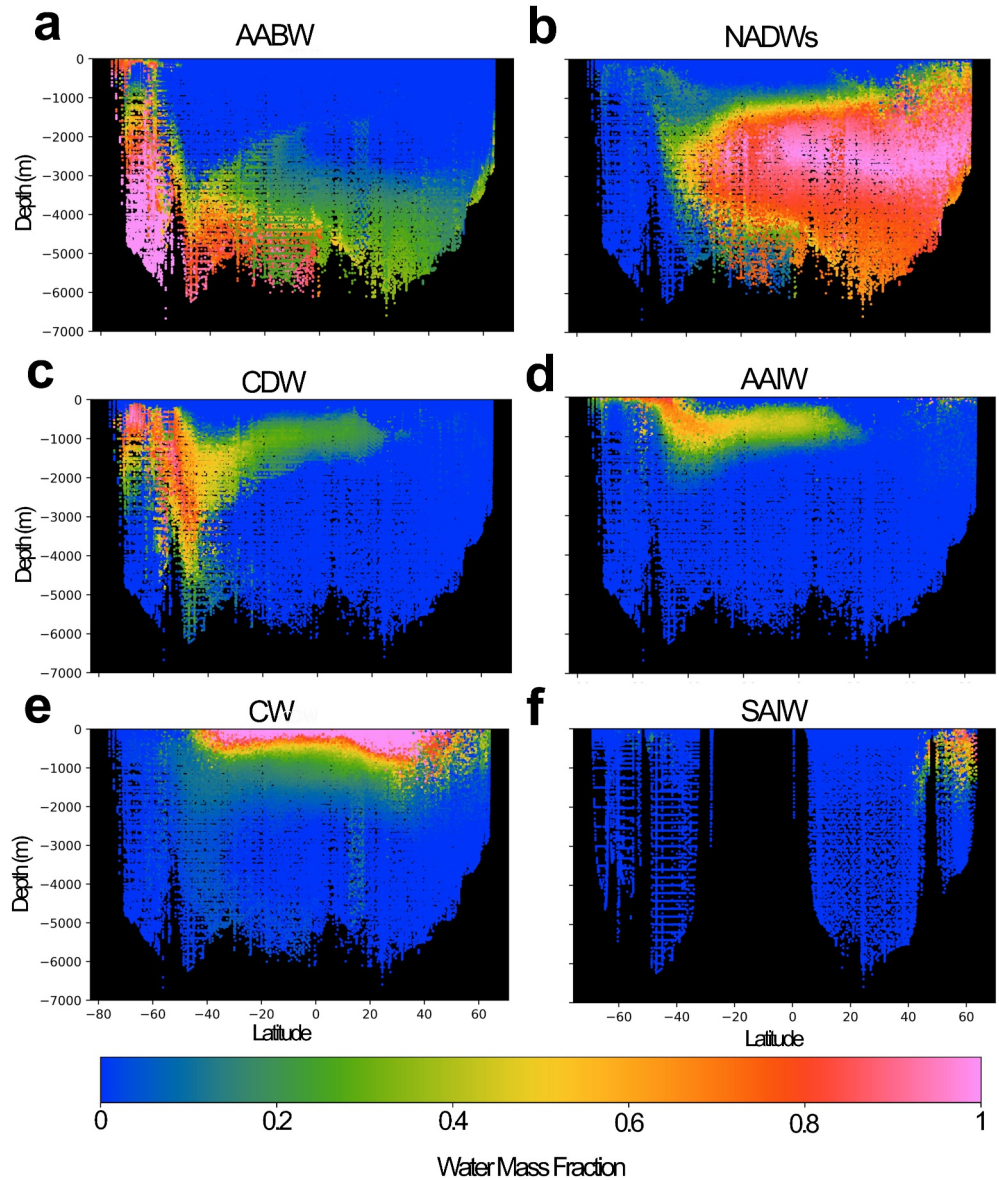

Figure 3. Zonal sections of water mass distribution in the Atlantic Ocean derived from Optimum Multiparameter (OMP) analysis of GLODAP data. Each panel shows a scatter plot of all available hydrographic profiles across the basin, colored by the proportion of a given water mass: (a) Antarctic Bottom Water, (b) North Atlantic Deep Waters, (c) Circumpolar Deep Water, (d) Antarctic Intermediate Water, (e) Central Water, and (f) Sub-Arctic Intermediate Water. Given the zonal asymmetry in Sub-Arctic Intermediate Water distribution, only data from longitudes 45–60°W are shown. Fractional contributions range from 0 (absence) to 1 (pure source water), highlighting vertical and meridional structure in the Atlantic water column.

the Labrador and Nordic Seas and export via the deep western boundary current system (Le Bras, 2023; Rhein et al., 2015; Talley, 2011; Toole et al., 2011). CDW dominates the deep Southern Ocean and appears in the abyssal Atlantic as a southward-flowing, nutrient-rich mixture that shoals toward Antarctica along tilted isopycnals associated with Southern Ocean upwelling (Figure 3c; Marshall & Speer, 2012). AAIW and SAIW are centered at intermediate depths, with AAIW subduction visible near the Polar Front (Figure 3d), while CW occupies the upper ocean in the subtropics (Figure 3e). In this analysis, the CW class also encompasses weakly stratified mode waters with similar thermohaline properties (Liu & Tanhua, 2021).

These spatial patterns are reflected in T-S space (Figure 2b), where CDW and AABW occupy the cold, saline end of the distribution, NADW spans intermediate temperatures and high salinities, intermediate waters occur at

lower salinities, and CW dominates the warm, saline surface regime. While a hard classification based on the dominant fraction clearly delineates major water masses, it can obscure transitions in regions of active mixing. To better represent continuous mixing, fractional compositions are visualized using an RGB blending approach (Figure 2c), revealing smooth gradients and transitional regimes, such as the gradual transition from NADW to AABW in the deep western South Atlantic (Memery et al., 2000).

Together, these results provide a physically consistent, tracer-informed reference classification that is used to train and evaluate the machine-learning model in the following sections.

3.2. Model Performance and Validation

Having established the OMP-derived reference structure of Atlantic water masses, we now assess how accurately this structure can be reproduced and extended using a machine-learning approach. Using the out-of-distribution testing approach outlined in Section 2.2, we first evaluate model performance using the full set of thermohaline, positional, and biogeochemical features. The model achieves strong agreement between predicted and OMP-derived water mass fractions, with coefficients of determination typically exceeding $R^2 \sim 0.9$ across most regions when trained on the full set of thermohaline, positional, and biogeochemical features (see Figure S5 in Supporting Information S1). The resulting model is therefore able to classify Atlantic water masses in complete biogeochemical data sets with substantially fewer requirements than OMP analysis; in particular, there is no need to define source water properties or prescribe parameter weights.

More importantly, further testing shows that the RF model ensemble retains strong performance even in the absence of biogeochemical tracers as input features. When trained using only temperature, salinity, and positional information, the model typically explains ~80–95% of the variance in OMP-derived water mass fractions across much of the Atlantic basin, with only moderate regional reductions in skill (Figure 4 and Figure S6 in Supporting Information S1). This demonstrates that robust, basin-scale water mass classification can be achieved using thermohaline and positional information alone.

To assess which input features are fundamental for skillful classification, we performed a set of feature-exclusion experiments (see Figure S6 in Supporting Information S1). In addition to the two aforementioned model set-ups, we also trained and validated two additional models following the same spatial holdout approach: one using only positional information, and one using only temperature and salinity. Here, the positional-only model serves as a spatial interpolation baseline, learning geographically coherent patterns rather than physically based water mass structure. The temperature-salinity only model shows substantially reduced skill in the subpolar North Atlantic, where attention to the individual water mass R^2 contributions reveals that it struggles to distinguish between SAIW and uNADW. It is likely that, in this region, additional positional or biogeochemical information is required to separate parcels that occupy overlapping ranges in thermohaline space but have different upstream pathways and ventilation histories. Conversely, the positional feature only model performs poorly in mid-latitudes and polar regions, where sharp temperature-salinity gradients associated with mode water formation, deep convection, and strong fronts (e.g., in the Labrador and Irminger Seas) determine water mass structure that geographic coordinates alone cannot capture. These results emphasize that, while positional context helps disambiguate similar source waters, reliable classification in dynamically active regions requires explicit thermohaline information. Overall, the exclusion study confirms that combining conservative temperature, absolute salinity, and positional features is sufficient to reproduce most of the skill of the full model, and that biogeochemical tracers further enhance performance only in specific regimes (primarily the Arctic and Antarctic shelves).

Whilst the thermohaline-positional model exhibits generally good predictive skill, it shows a degree of variability across the basin. Lower R^2 values tend to occur in regions where (a) water mass structure is highly variable across small spatial scales (e.g., strong mixing zones, such as the Drake Passage where $R^2 = 0.78$), or (b) there is undersampling in the GLODAP data set, limiting the statistical representation available for training (such as the box due south of South Africa where $R^2 = 0.67$, or the box in the Arctic where $R^2 = 0.70$). Near the Antarctic shelf, model skill is also reduced despite relatively dense sampling (e.g., $R^2 = 0.60$ in the Weddell Sea). This likely reflects a mismatch between our chosen open-ocean end-members and the local shelf waters which are modified by relatively strong non-conservative buoyancy fluxes (e.g., Dense Shelf Water, Ice Shelf Water, or Winter Water, which are not included as end-members for simplicity here, but in principle could be—see Lanham,

10.1029/2025JH001182

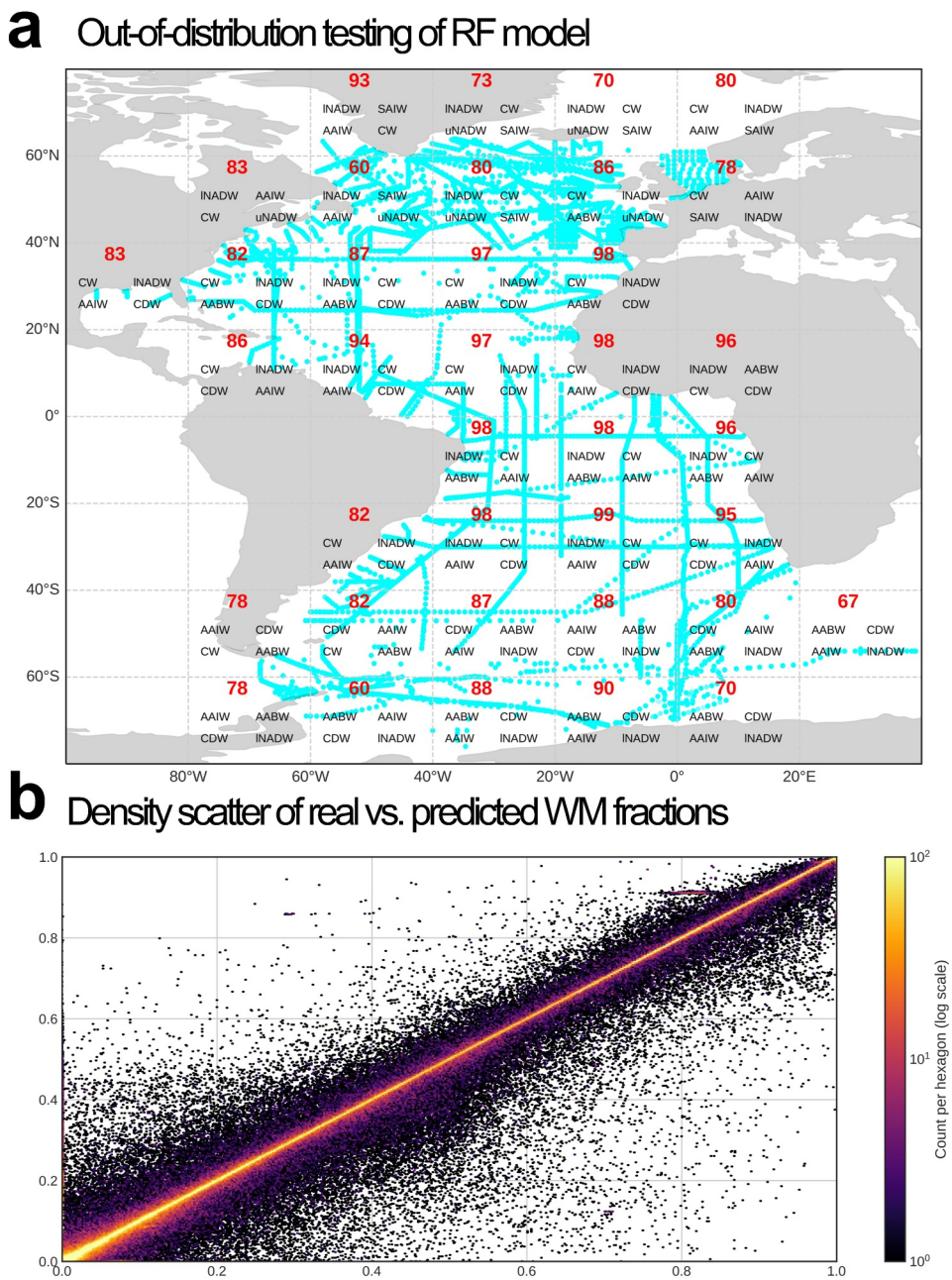

Figure 4. Verification of the RF approach to classifying water masses in the absence of biogeochemical tracers. Panel (a) shows an out-of-distribution testing scenario. The Atlantic Ocean is divided into boxes and used to geographically bin all the GLODAP data used in this study, shown by blue dots. Profiles that fall within a given box are separated from the data in the remaining boxes and withheld from training. The RF ensemble is then trained on the data from the remaining boxes and tested on the data from the original box. The red number represents the R^2 value associated with this testing (*100). Note that we exclude any water masses from the calculation in regions in which those same water masses were effectively absent ($n > 1$, where n is a single measurement in which the water mass fraction >0.1). The four water masses with the highest relative fraction in each box are indicated. Panel (b) shows a bivariate histogram of all the OMP-derived water mass fractions (x-axis), plotted against the model-derived water mass fraction prediction (y-axis).

Mazloff, et al. (2025)). In such regions, the OMP framework forces mixtures of the available end-members to represent water types that are not well spanned in tracer space, likely leading to noisier target fractions and lower apparent skill in the RF predictions.

By contrast, in interior or well-sampled regions, the RF performs strongly, with predictions tightly aligned with OMP-derived fractions. This means that skillful water mass classification in the Atlantic Ocean can be achieved in observational data sources that only measure temperature and salinity. This greatly increases the volume of classifiable data, enabling water mass classification at much finer spatial and temporal resolution, particularly thanks to the widespread coverage of Argo floats (Roemmich et al., 2009) (see Text S1, Figure S1 in Supporting Information S1 for more details).

The RF ensemble is also able to classify water masses in ocean models that do not have a biogeochemical component. Here we demonstrate this functionality by applying the RF model ensemble to the monthly gridded data-assimilating ECCOv4r4 state estimate to produce a fully gridded water mass product at monthly resolution. This enables a preliminary exploration into how the distribution of water masses in the Atlantic Ocean has evolved in recent decades.

3.3. Basin-Wide Extension of Classification Between 1992 and 2018

3.3.1. Water Mass Climatology

Figure 5 shows the climatological distribution of major Atlantic Ocean water masses at selected depth levels, diagnosed by applying the RF ensemble to the ECCO state estimate. The resulting three-dimensional fields reproduce the expected large-scale structure and pathways of Atlantic water masses, consistent with hydrographic observations and the OMP-based reference classification (See Figure S7 in Supporting Information S1 for the RF-ECCO equivalent to Figure 3).

At abyssal depths (4,000 m), AABW dominates the South Atlantic (Figure 5a), with high fractions near the Antarctic margin in the Weddell Gyre that decrease northward, in agreement with observations (Silvano et al., 2023). The influence of bathymetry on bottom water pathways is clearly evident, with northward transport along the western boundary through the Argentine and Brazil basins and restricted exchange across major topographic features such as the Walvis Ridge and the Mid-Atlantic Ridge (Durrieu De Madron & Weatherly, 1994; Sandoval & Weatherly, 2001; Solodoch et al., 2022). At 2,500 m, AABW is largely confined to the Weddell Sea sector, reflecting downslope export of dense shelf waters formed in coastal polynyas (Amblas & Dowdeswell, 2018).

NADW exhibits a layered vertical structure, with a strong western boundary signal at 1,500 m extending from the subpolar North Atlantic to ~20°S (Figure 5e), and a broader, more diffuse distribution at 3,000 m that fills much of the basin interior and extends to ~40°S (Figure 5d), consistent with observations (Ferreira & Kerr, 2017; Johnson, 2008; Liu & Tanhua, 2021; Miron et al., 2022).

CDW is apparent at both 1,500 and 2,500 m in the South Atlantic between 40°S and 60°S (Figures 5c and 5f), forming a zonal band that separates NADW to the north from AABW to the south. This structure reflects Southern Ocean wind forcing and associated upwelling linked to Ekman divergence (Lumpkin & Speer, 2007; Marshall & Speer, 2012; Talley, 2013), with onshore encroachment limited by dense water formation in the Weddell Sea (Davis et al., 2022; Lanham, Mazloff, et al., 2025; Tamsitt et al., 2021). At intermediate depths (500 m), AAIW forms a zonal band between 30°S and 50°S and spreads northward beyond the equator (Figure 5i), while SAIW is confined to the western boundary of the subpolar North Atlantic (Figure 5g; Arhan, 1990; Talley, 1996). CW dominates the upper ocean at 100 m in the low-to-mid latitudes (Figure 5h).

Overall, these climatological fields demonstrate that the RF–OMP hybrid approach produces physically realistic, basin-scale water mass distributions when applied to gridded ocean state estimates. In the following section, we use this framework to examine temporal variability in Atlantic water mass structure over the past three decades.

3.3.2. Water Mass Temporal Variability

Application of the RF model ensemble to high-resolution data raises the possibility of analyzing temporal variability in the Atlantic Ocean water mass structure, albeit with the caveat that we keep the properties of the SWT end member fixed in time. This means that temporal variability in source water properties can impact the perceived variability in water mass distribution (discussed further in Section 4). We make this choice given that the focus of this work is primarily to showcase the methodology, however, we note that it is possible to relax this

10.1029/2025JH001182

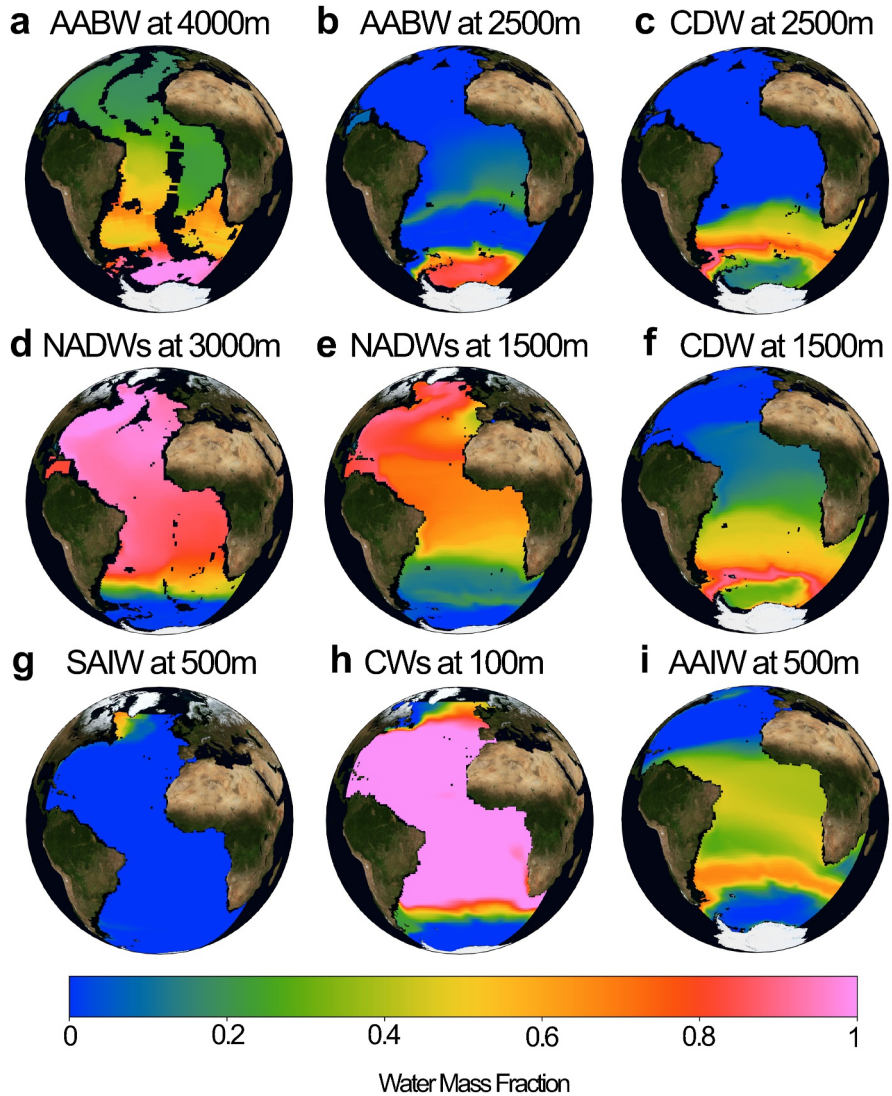

Figure 5. Climatological (1992–2018) water mass distributions in the Atlantic Ocean within the ECCO model. The RF ensemble is applied to the data-assimilating ECCO ocean state estimate in the absence of biogeochemical tracers and averaged across the 30-year window to produce a "mean-state" estimate of the 3-dimensional Atlantic Ocean water mass structure. (a) Antarctic Bottom Water at 4,000 m; (b) Antarctic Bottom Water at 2,500 m. (c) Circumpolar Deep Water at 2,500 m; (d) North Atlantic Deep Waters (sum of upper and lower components) at 3,000 m; (e) North Atlantic Deep Waters at 1,500 m; (f) Circumpolar Deep Water at 1,500 m; (g) Sub-Arctic Intermediate Water at 500 m; (h) Atlantic Central Water at 100 m; (i) Antarctic Intermediate Water at 500 m.

assumption if end members are allowed to vary in time in accordance with changes in the properties of the source water region (see Lanham, Purkey, et al., 2025).

Figures 6a and 6b show the change in the layer thickness of CW and AAIW over the ECCO period (anomaly, 2013–2018 average minus 1992–1997 average), according to the RF ensemble applied to ECCO. The equivalent time series of the total volume of CW south of 30°S and AAIW are shown in Figures 6c and 6d, respectively. CW undergoes a noticeable thickening south of 30°S, with both layer thickness and integrated volume increasing over time. This expansion of the warm, saline upper layer may reflect enhanced surface warming and increased subtropical stratification in the Southern Hemisphere, potentially linked to intensified surface buoyancy forcing and changes in wind-driven subduction. This is consistent with a number of studies that note a recent increase in subduction rates and accumulation of mode water volume (Gao et al., 2018; Kolodziejczyk et al., 2019; Qu et al., 2020; Portela et al., 2020; Z. Li et al., 2023). In contrast, AAIW shows a declining trend in both layer

10.1029/2025JH001182

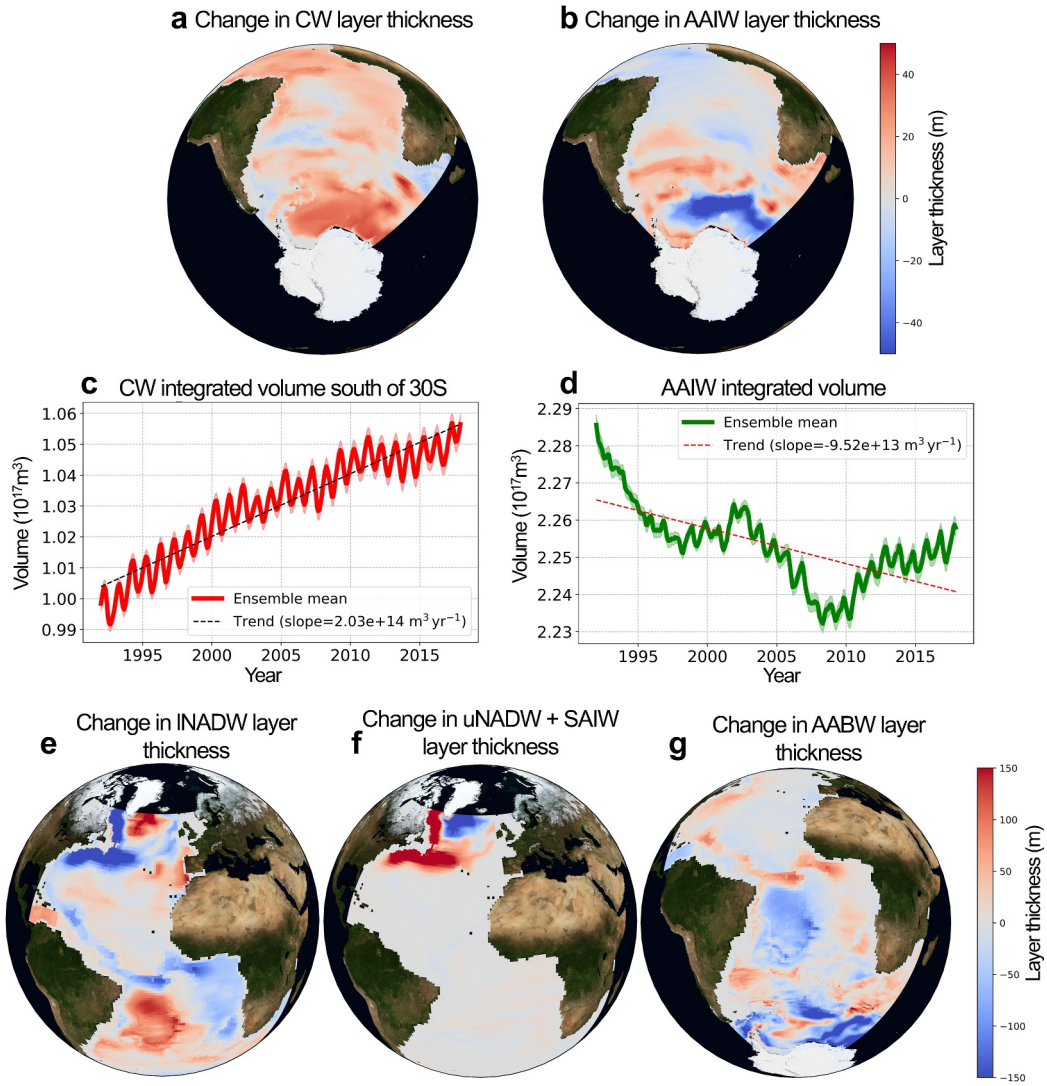

Figure 6. Changes in water mass distributions in the Atlantic Ocean within the ECCO model, obtained via application of the RF ensemble in the absence of biogeochemical tracers. Panels (a, b, e–g) show the change in integrated full-column water mass layer thicknesses across the 30 years ECCO run. The change in thickness is calculated as the mean thickness of the last 5 years, minus the mean of the first 5 years. Panels (c, d) show the integrated volume timeseries of Atlantic Central Water south of 30°S and Antarctic Intermediate Water across the whole basin. Shading indicates ±2 standard deviations across the 5 k-fold ensemble member predictions.

thickness and total volume (Figures 6c and 6d), consistent with reduced formation or shoaling of intermediate isopycnals. This is also consistent with observed trends in the Southern Ocean in recent decades (Z. Li et al., 2023; Portela et al., 2020). Together, these trends suggest a deepening and thickening of the upper ocean and a contraction of the intermediate layer, indicative of a more strongly stratified Southern Ocean in recent decades (Haumann et al., 2016).

Figures 6e and 6f show the change in the layer thickness of lNADW and uNADW + SAIW during the same period. uNADW and SAIW are grouped because of a similar pattern of anomalies that appear to oppose the changes in lNADW in the subpolar North Atlantic. The lNADW layer thickens in the Irminger Sea but thins farther south along the western boundary. In contrast, the upper NADW + SAIW layer thins in the Irminger Sea sector and thickens near the Labrador Sea outflow, which is possibly consistent with the recent increase in deep convection observed in the Labrador Sea (Yashayaev, 2024; Yashayaev & Loder, 2016). Increased Labrador Sea Water volume is also found when the RF ensemble is trained directly on the Liu and Tanhua (2021) classification

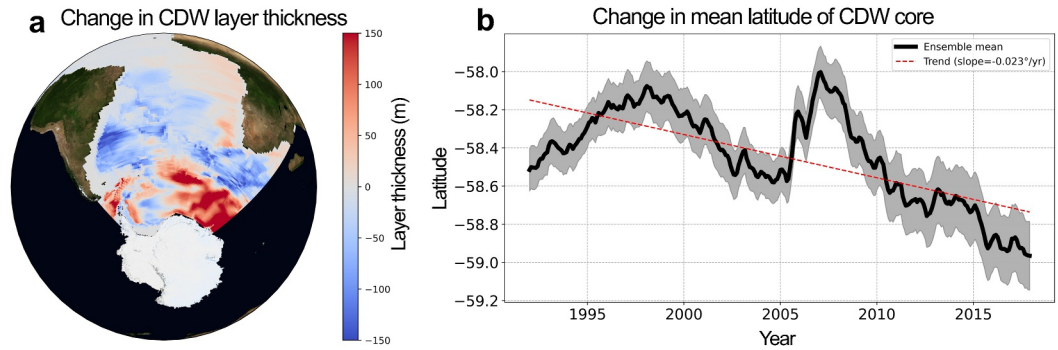

Figure 7. Changes in Circumpolar Deep Water distribution in the Atlantic Ocean within the ECCO model, obtained via application of the RF ensemble in the absence of biogeochemical tracers. Panel (a) shows the change in integrated layer thickness of Circumpolar Deep Water. The change in thickness is calculated as the mean thickness of the last 5 years, minus the mean of the first 5 years. Panel (b) shows the change in the mean latitude of the core of Circumpolar Deep Water, defined as all grid points in which the relative concentration exceeds 75%. Shading indicates ±2 standard deviations across the 5 k-fold ensemble member predictions.

(see Figure S4 in Supporting Information S1). By contrast, AABW shows significant layer contraction across most of the basin (Figure 6g), consistent with the wide array of observations that show a warming, freshening and contraction of AABW in recent decades (Gunn et al., 2023; Johnson & Purkey, 2024; Purkey & Johnson, 2010, 2012; Silvano et al., 2023).

Figure 7a shows the change in integrated water mass layer thickness over the 30-year ECCO period for CDW, computed over the full water column. CDW undergoes a pronounced expansion near the Antarctic continent, with modest reductions on the equatorward side of the ACC. This pattern indicates a basin-scale redistribution of CDW toward the Antarctic margin. When the analysis is restricted to the upper 2000 m (Figure S8 in Supporting Information S1), the CDW expansion is concentrated clearly along the Antarctic continental shelf, highlighting the accumulation of CDW in the depth range most relevant for shelf–ocean heat exchange. The increase in CDW is largely compensated by a reduction in AABW fraction, which at these depths primarily represents recently ventilated dense shelf waters and the upper expression of newly formed AABW cascading from the Antarctic margin rather than the abyssal AABW core.

The spatial structure of the CDW anomalies is plausibly consistent with a poleward shift of the upwelling branch of the circulation, with relative CDW fractions decreasing equatorward of the upwelling zone and increasing on its poleward flank. This interpretation is supported by Figure 7b, which shows that the mean latitude of the CDW core (defined here as regions exceeding 75% CDW fraction) migrates poleward by approximately half a degree over the 30-year period. Similar behavior is recovered when the RF ensemble is trained on the alternative OMP classification of Liu and Tanhua (2021) (Figure S4 in Supporting Information S1), and is consistent with observational analyses of CDW near the Antarctic shelf (Herraiz-Borreguero & Naveira Garabato, 2022; Lanham, Purkey, et al., 2025; Schmidtko et al., 2014) as well as model projections of future Southern Ocean change Q. Li et al. (2023). We note, however, that the inferred magnitude of this shift is associated with substantial uncertainty (gray shading), reflecting spread across the k-fold model ensemble.

The explicit aim of the analysis presented here is not to explain or uncover new phenomenology. Rather, it is to demonstrate that the OMP-RF hybrid approach is consistent with current understanding of Atlantic water mass distributions and that it provides a practical tool for studying AMOC variability, allowing for accurate estimates of heat, carbon, and nutrient budgets.

4. Discussion and Conclusions

We present a hybrid method for water mass classification of the Atlantic Ocean, combining the tracer constraints of OMP analysis with the scalability and ease-of-application of machine learning. This approach leverages a RF ensemble trained on a biogeochemical tracer-informed OMP-derived data set to classify water masses using only temperature, salinity, and positional information. This enables the identification of water masses across a larger

range of data at higher spatio-temporal resolution, including Argo profiles, CTD-only observations, and numerical model outputs without a biogeochemical component.

Applying this approach to the ECCO state estimate, we produce a monthly gridded water mass product for the Atlantic Ocean over the period 1992–2018. This product reveals climatological patterns that align with established understanding of Atlantic Ocean structure, while also enabling an assessment of decadal-scale variability. We observe a consistent contraction in the thickness and volume of AABW in the Weddell Sea and abyssal South Atlantic. This finding is consistent with previous observational studies documenting AABW warming, freshening, and contraction, and may indicate a weakening of the abyssal overturning circulation (Biló et al., 2024). A thinning AABW layer reduces the capacity of the deep ocean to ventilate and store heat and carbon, with implications for global climate regulation (Akhoudas et al., 2021; Gunn et al., 2023; Huhn et al., 2013; Lago & England, 2019; Nissen et al., 2022). In the North Atlantic, we detect opposing trends in upper and lower NADW layers: a thickening of upper NADW (and SAIW) near the Labrador Sea outflow region, and a thinning of lower NADW in the same area, with the reverse pattern in the Irminger Sea. This is at least partially consistent with recent increases in ventilation in the Labrador Sea (Yashayaev, 2024; Yashayaev & Loder, 2016).

In the mid-depth Southern Ocean, we find preliminary evidence for a southward shift in the core of CDW, with expansion on the poleward side of the upwelling zone and a minor reduction to the north. This pattern is plausibly consistent with a latitudinal reorganisation of upwelling, possibly related to a slowdown in AABW formation (Q. Li et al., 2023) or wind-driven shifts in the ACC. In particular, the observed poleward shift of Southern Hemisphere westerlies (Deng et al., 2022; Thomas et al., 2015) may have enhanced Ekman divergence closer to the Antarctic continent, repositioning CDW upwelling zones. Alternatively, it is possible that the observed strengthening in these winds could drive poleward CDW migration via a strengthened eddy response (Gille, 2014; Goyal et al., 2021; O'Connor et al., 2021). Such changes have important implications for heat transport to ice shelf fronts and may modulate basal melting rates (Nakayama et al., 2018; Pritchard et al., 2012; Schodlok et al., 2016). In the upper and intermediate ocean, we observe a thickening of warm central waters and a contraction of AAIW, particularly in the Southern Hemisphere. These results could reflect increased stratification in the upper ocean driven by surface warming and strengthened buoyancy fluxes, or the wind-driven deepening of the surface mixed layer linked with increasingly strong westerlies over the Southern Ocean (Deng et al., 2022).

When interpreting temporal changes in diagnosed water mass thickness and volume, it is important to recognize that our analysis assumes temporally fixed source water properties. As a result, changes in inferred water mass extent may reflect a combination of (a) genuine redistribution of water masses within the basin and (b) temporal evolution of source water properties in their formation regions. In particular, documented warming and freshening of AABW and NADW over recent decades may cause portions of these waters to project less strongly onto fixed end members, leading to an apparent reduction in fractional extent even in the absence of a purely dynamical contraction.

This limitation should be borne in mind when interpreting trends. Nevertheless, several aspects of our results suggest that the diagnosed changes are not solely artifacts of evolving source properties. The spatial structure of the anomalies is physically coherent, aligns with known circulation pathways, and is consistent with independent observational and modeling studies documenting abyssal warming, freshening, and contraction. Moreover, the same qualitative patterns emerge when the RF ensemble is trained on an alternative OMP classification Liu and Tanhua (2021), lending confidence that the inferred changes reflect basin-scale reorganization rather than local end-member drift alone. Finally, sensitivity tests presented in Lanham, Purkey, et al. (2025), conducted in the Southern Ocean, show that allowing source water properties to vary in time does not substantially alter the sign or large-scale structure of diagnosed water mass trends. While regionally distinct, this result suggests that fixed end-member assumptions are unlikely to dominate basin-scale trend detection, supporting the robustness of the patterns identified here.

Accounting explicitly for time-varying source water properties through adaptive or time-dependent end members is a promising avenue for future work and has been explored in regional and basin-scale studies (e.g., Lanham, Purkey, et al. (2025) and Oglethorpe et al. (2026)). Implementing such an approach here would require additional assumptions and lies beyond the scope of this primarily methodological study.

More generally, it should be noted that the definition of source water properties is not uniquely prescribed: end members must ultimately be selected based on a combination of physical understanding, observational evidence,

and methodological judgment. Different but equally defensible choices of end-member properties can therefore lead to quantitatively different water mass fractions, even when applied to the same underlying data.

Here we deliberately adopt an idealized representation of the Atlantic Ocean using seven basin-scale water masses, acknowledging that this is a simplification of the full diversity of water mass structure observed in regional and process-focused studies. The number of water masses that can be resolved in this framework is set by the OMP end-member definition rather than by the machine-learning model itself. While the RF can readily accommodate a larger number of output classes, increasing the number of basin-scale end members in this case leads to an underdetermined OMP system, in which solutions are no longer unique and depend more strongly on parameter weights, priors, and additional regularization. Such formulations can be appropriate when carefully constrained; however, because the primary aim of this study is to demonstrate a basin-wide, gridded extension of water mass classification, we adopt a reduced set of physically distinct water masses to prioritize methodological clarity, robustness, and spatial coherence.

Overall, the RF-based approach provides a practical and scalable framework for extending water mass classification to the rapidly growing volume of high-resolution physical oceanographic data. This study demonstrates that robust basin-scale water mass classification can be achieved using temperature and salinity alone, without reliance on biogeochemical tracers. This substantially broadens the range of observational and modeling datasets to which water mass analysis can be applied, including Argo floats, gliders, and ocean state estimates that lack a biogeochemical component.

A growing body of recent work has already adopted this hybrid OMP–RF framework. For example, Lanham, Purkey et al. (2025) apply it to Southern Ocean Argo float data, while Oglethorpe et al. (2026) use it with ice-tethered profilers in the Arctic; both studies employ adaptive end-member strategies to assess multi-year changes in water mass structure. Dias et al. (2025) apply the method to investigate regions of water mass mixing in the Atlantic Ocean on palaeo-oceanographic timescales. Collectively, these applications highlight the versatility of the hybrid OMP–RF approach, and demonstrate its utility for studying both the mean state and evolving structure of the global ocean using widely available physical observations.

Conflict of Interest

The authors declare no conflicts of interest relevant to this study.

Availability Statement

All ship section data used in this study were obtained via the GLODAPv2.2023 merged and adjusted product (Lauvset et al., 2024), https://doi.org/10.5194/essd-16-2047-2024. This is available at https://glodap.info. Water mass fractions calculated by Liu and Tanhua (2021) are available at https://doi.org/10.5194/os-17-463-2021. ECCO Version 4 Release 4 (V4r4) temperature and salinity fields spanning 1992–2018 were obtained from the NASA PO.DAAC archive (https://podaac.jpl.nasa.gov/ECCO), https://doi.org/10.5067/ECG5D-OTS44. The code used in this study is archived on Zenodo (Lanham, 2026) and available at: https://doi.org/10.5281/zenodo.19451466.

Acknowledgments
J.L. acknowledges EPSRC doctoral training grant EP/T51780X/1. K.S. was supported by grants from Office of Naval Research (N00014-25-1-2183 and N00014-20-1-2023). L.C. received support through Schmidt Sciences, LCC, and the Advanced Research and Invention Agency (ARIA) under grant G130701. A.M. acknowledges support from US ONR (Grant N00014-22-1-2082) and UK ARIA.

References

Akhoudas, C. H., Sallée, J.-B., Haumann, F. A., Meredith, M. P., Garabato, A. N., Reverdin, G., et al. (2021). Ventilation of the abyss in the Atlantic sector of the Southern Ocean. *Scientific Reports*, *11*(1), 6760. https://doi.org/10.1038/s41598-021-86043-2

Amblas, D., & Dowdeswell, J. A. (2018). Physiographic influences on dense shelf-water cascading down the Antarctic continental slope. *Earth-Science Reviews*, *185*, 887–900. https://doi.org/10.1016/j.earscirev.2018.07.014

Anilkumar, N., Jena, B., George, J. V., P, S., S, K., & Ravichandran, M. (2021). Recent freshening, warming, and contraction of the Antarctic bottom water in the Indian sector of the Southern Ocean. *Frontiers in Marine Science*, *8*, 730630. https://doi.org/10.3389/fmars.2021.730630

Arhan, M. (1990). The north Atlantic Current and subarctic intermediate water. *Journal of Marine Research*, *48*(1), 109–144. https://doi.org/10.1357/002224090784984605

Biló, T. C., Perez, R. C., Dong, S., Johns, W., & Kanzow, T. (2024). Weakening of the Atlantic Meridional Overturning Circulation abyssal limb in the North Atlantic. *Nature Geoscience*, *17*(5), 419–425. https://doi.org/10.1038/s41561-024-01422-4

Broecker, W. S. (1974). "NO", a conservative water-mass tracer. Earth and Planetary Science Letters. *Earth and Planetary Science Letters*, *23*(1), 100–107. https://doi.org/10.1016/0012-821X(74)90036-3

Caesar, L., McCarthy, G. D., Thornalley, D. J. R., Cahill, N., & Rahmstorf, S. (2021). Current Atlantic Meridional Overturning Circulation weakest in last millennium. *Nature Geoscience*, *14*(3), 118–120. https://doi.org/10.1038/s41561-021-00699-z

Cimoli, L., Gebbie, G., Purkey, S. G., & Smethie, W. M. (2023). Annually resolved propagation of CFCs and SF_6 in the Global Ocean over eight decades. *Journal of Geophysical Research: Oceans*, *128*(3), e2022JC019337. https://doi.org/10.1029/2022JC019337

Cimoli, L., Mashayek, A., Johnson, H. L., Marshall, D. P., Naveira Garabato, A. C., Whalen, C. B., et al. (2023). Significance of diapycnal mixing within the Atlantic meridional overturning circulation. *AGU Advances*, *4*(2), e2022AV000800. https://doi.org/10.1029/2022AV000800

Davis, P. E. D., Jenkins, A., Nicholls, K. W., Dutrieux, P., Schröder, M., Janout, M. A., et al. (2022). Observations of modified warm deep water beneath Ronne Ice Shelf, Antarctica, from an autonomous underwater vehicle. *Journal of Geophysical Research: Oceans*, *127*(11), e2022JC019103. https://doi.org/10.1029/2022JC019103

Deng, K., Azorin-Molina, C., Yang, S., Hu, C., Zhang, G., Minola, L., & Chen, D. (2022). Changes of Southern Hemisphere westerlies in the future warming climate. *Atmospheric Research*, *270*, 106040. https://doi.org/10.1016/j.atmosres.2022.106040

Dias, B. B., Chiessi, C. M., Piotrowski, A. M., Campos, M. C., Jena, P. S., Ballalai, J. M., et al. (2025). Reduced penetration of northern-sourced waters into the south Atlantic during the last interglacial relative to the Holocene. *Paleoceanography and Paleoclimatology*, *40*(2), e2024PA004955. https://doi.org/10.1029/2024PA004955

Ditlevsen, P., & Ditlevsen, S. (2023). Warning of a forthcoming collapse of the Atlantic meridional overturning circulation. *Nature Communications*, *14*(1), 4254. https://doi.org/10.1038/s41467-023-39810-w

Durrieu De Madron, X., & Weatherly, G. (1994). Circulation, transport and bottom boundary layers of the deep currents in the Brazil Basin. *Journal of Marine Research*, *52*(4), 583–638. https://doi.org/10.1357/0022240943076975

ECCO Consortium, Fukumori, I., Wang, O., Fenty, I., Forget, G., Heimbach, P., & Ponte, R. M. (2021). Synopsis of the ECCO central production global Ocean and sea-ice State estimate, version 4 release 4 (Tech. Rep.). *Zenodo*. https://doi.org/10.5281/ZENODO.4533349

Ellison, E., Cimoli, L., & Mashayek, A. (2023). Multi-time scale control of Southern Ocean diapycnal mixing over Atlantic tracer budgets. *Climate Dynamics*, *60*(9–10), 3039–3050. https://doi.org/10.1007/s00382-022-06428-5

Ferreira, M. L. D. C., & Kerr, R. (2017). Source water distribution and quantification of North Atlantic Deep Water and Antarctic Bottom Water in the Atlantic Ocean. *Progress in Oceanography*, *153*, 66–83. https://doi.org/10.1016/j.pocean.2017.04.003

Forget, G., Campin, J.-M., Heimbach, P., Hill, C. N., Ponte, R. M., & Wunsch, C. (2015). ECCO version 4: An integrated framework for non-linear inverse modeling and global ocean state estimation. *Geoscientific Model Development*, *8*(10), 3071–3104. https://doi.org/10.5194/gmd-8-3071-2015

Gao, L., Rintoul, S. R., & Yu, W. (2018). Recent wind-driven change in Subantarctic Mode Water and its impact on ocean heat storage. *Nature Climate Change*, *8*(1), 58–63. https://doi.org/10.1038/s41558-017-0022-8

Gille, S. T. (2014). Meridional displacement of the Antarctic Circumpolar Current. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *372*(2019), 20130273. https://doi.org/10.1098/rsta.2013.0273

Goyal, R., Sen Gupta, A., Jucker, M., & England, M. H. (2021). Historical and projected changes in the Southern hemisphere surface westerlies. *Geophysical Research Letters*, *48*(4), e2020GL090849. https://doi.org/10.1029/2020GL090849

Grinsztajn, L., Oyallon, E., & Varoquaux, G. (2022). Why do tree-based models still outperform deep learning on tabular data? *arXiv*. https://doi.org/10.48550/ARXIV.2207.08815

Gunn, K. L., Rintoul, S. R., England, M. H., & Bowen, M. M. (2023). Recent reduced abyssal overturning and ventilation in the Australian Antarctic Basin. *Nature Climate Change*, *13*(6), 537–544. https://doi.org/10.1038/s41558-023-01667-8

Haumann, F. A., Gruber, N., Münnich, M., Frenger, I., & Kern, S. (2016). Sea-ice transport driving Southern Ocean salinity and its recent trends. *Nature*, *537*(7618), 89–92. https://doi.org/10.1038/nature19101

Herraiz-Borreguero, L., & Naveira Garabato, A. C. (2022). Poleward shift of circumpolar deep water threatens the East Antarctic Ice sheet. *Nature Climate Change*, *12*(8), 728–734. https://doi.org/10.1038/s41558-022-01424-3

Huhn, O., Rhein, M., Hoppema, M., & Van Heuven, S. (2013). Decline of deep and bottom water ventilation and slowing down of anthropogenic carbon storage in the Weddell Sea, 1984–2011. *Deep Sea Research Part I: Oceanographic Research Papers*, *76*, 66–84. https://doi.org/10.1016/j.dsr.2013.01.005

Jacobs, S., Helmer, H., Doake, C. S. M., Jenkins, A., & Frolich, R. M. (1992). Melting of ice shelves and the mass balance of Antarctica. *Journal of Glaciology*, *38*(130), 375–387. https://doi.org/10.3189/S0022143000002252

Johnson, G. C. (2008). Quantifying Antarctic bottom water and north Atlantic deep water volumes. *Journal of Geophysical Research*, *113*(5). https://doi.org/10.1029/2007JC004477

Johnson, G. C., & Purkey, S. G. (2024). Refined estimates of global Ocean deep and abyssal decadal warming trends. *Geophysical Research Letters*, *51*(18), e2024GL111229. https://doi.org/10.1029/2024GL111229

Karstensen, J., & Tomczak, M. (1998). Age determination of mixed water masses using CFC and oxygen data. *Journal of Geophysical Research*, *103*(C9), 18599–18609. https://doi.org/10.1029/98JC00889

Kobayashi, T. (2018). Rapid volume reduction in Antarctic Bottom Water off the Adélie/George V Land coast observed by deep floats. *Deep Sea Research Part I: Oceanographic Research Papers*, *140*, 95–117. https://doi.org/10.1016/j.dsr.2018.07.014

Kolodziejczyk, N., Llovel, W., & Portela, E. (2019). Interannual variability of upper Ocean water masses as inferred from Argo Array. *Journal of Geophysical Research: Oceans*, *124*(8), 6067–6085. https://doi.org/10.1029/2018JC014866

Lago, V., & England, M. H. (2019). Projected slowdown of Antarctic bottom water Formation in response to amplified Meltwater Contributions. *Journal of Climate*, *32*(19), 6319–6335. https://doi.org/10.1175/JCLI-D-18-0622.1

Lanham, J. (2026). Code for: Basin-wide Atlantic Ocean water mass classification and climatic variability from machine learning (Version 1.0.0). *Zenodo*. https://doi.org/10.5281/ZENODO.19451467

Lanham, J., Mazloff, M., Naveira Garabato, A. C., Siegert, M., & Mashayek, A. (2025). Seasonal regimes of warm Circumpolar Deep Water intrusion toward Antarctic ice shelves. *Communications Earth and Environment*, *6*(1), 168. https://doi.org/10.1038/s43247-025-02091-w

Lanham, J., Purkey, S., Srinivasan, K., Mazloff, M., Cimoli, L., & Mashayek, A. (2025). Observational evidence for a poleward migration of warm Circumpolar Deep Water towards Antarctica. *Review*. https://doi.org/10.21203/rs.3.rs-7021596/v1

Lauvset, S. K., Key, R. M., Olsen, A., van Heuven, S., Velo, A., Lin, X., et al. (2016). A new global interior ocean mapped climatology: The 1°×1° GLODAP version 2.

Lauvset, S. K., Lange, N., Tanhua, T., Bittig, H. C., Olsen, A., Kozyr, A., et al. (2024). The annual update GLODAPv2.2023: The global interior ocean biogeochemical data product. *Earth System Science Data*, *16*(4), 2047–2072. https://doi.org/10.5194/essd-16-2047-2024

Le Bras, I. A.-A. (2023). Labrador sea water spreading and the Atlantic meridional overturning circulation. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *381*(2262), 20220189. https://doi.org/10.1098/rsta.2022.0189

Li, Q., England, M. H., Hogg, A. M., Rintoul, S. R., & Morrison, A. K. (2023). Abyssal ocean overturning slowdown and warming driven by Antarctic meltwater. *Nature*, *615*(7954), 841–847. https://doi.org/10.1038/s41586-023-05762-w

Li, Z., England, M. H., & Groeskamp, S. (2023). Recent acceleration in global ocean heat accumulation by mode and intermediate waters. *Nature Communications*, *14*(1), 6888. https://doi.org/10.1038/s41467-023-42468-z

Liu, M., & Tanhua, T. (2021). Water masses in the Atlantic Ocean: Characteristics and distributions. *Ocean Science*, *17*(2), 463–486. https://doi.org/10.5194/os-17-463-2021

Lumpkin, R., & Speer, K. (2007). Global Ocean meridional overturning. *Journal of Physical Oceanography*, *37*(10), 2550–2562. https://doi.org/10.1175/JPO3130.1

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

Mashayek, A., Ferrari, R., Nikurashin, M., & Peltier, W. R. (2015). Influence of enhanced abyssal diapycnal mixing on stratification and the Ocean overturning circulation. *Journal of Physical Oceanography*, *45*(10), 2580–2597. https://doi.org/10.1175/JPO-D-15-0039.1

Mashayek, A., Reynard, N., Zhai, F., Srinivasan, K., Jelley, A., Naveira Garabato, A., & Caulfield, C. P. (2022). Deep Ocean learning of small Scale turbulence. *Geophysical Research Letters*, *49*(15), e2022GL098039. https://doi.org/10.1029/2022GL098039

Mémery, L., Arhan, M., Alvarez-Salgado, X., Messias, M.-J., Mercier, H., Castro, C., & Rios, A. (2000). The water masses along the western boundary of the south and equatorial Atlantic. *Progress in Oceanography*, *47*(1), 69–98. https://doi.org/10.1016/S0079-6611(00)00032-X

Miron, P., Beron-Vera, F. J., & Olascoaga, M. J. (2022). Transition paths of north Atlantic deep water. *Journal of Atmospheric and Oceanic Technology*, *39*(7), 959–971. https://doi.org/10.1175/JTECH-D-22-0022.1

Nakayama, Y., Menemenlis, D., Zhang, H., Schodlok, M., & Rignot, E. (2018). Origin of Circumpolar Deep Water intruding onto the Amundsen and Bellingshausen Sea continental shelves. *Nature Communications*, *9*(1), 3403. https://doi.org/10.1038/s41467-018-05813-1

Nissen, C., Timmermann, R., Hoppema, M., Gürses, Ã., & Hauck, J. (2022). Abruptly attenuated carbon sequestration with Weddell Sea dense waters by 2100. *Nature Communications*, *13*(1), 3402. https://doi.org/10.1038/s41467-022-30671-3

O'Connor, G. K., Steig, E. J., & Hakim, G. J. (2021). Strengthening Southern hemisphere westerlies and Amundsen Sea low deepening over the 20th century revealed by proxy–data assimilation. *Geophysical Research Letters*, *48*(24), e2021GL095999. https://doi.org/10.1029/2021GL095999

Oglethorpe, K., Lanham, J., Reiss, R. S., Boland, E. J. D., Naveira Garabato, A. C., Caulfield, C.-C. P., & Mashayek, A. (2026). Water masses of the Arctic from 40 years of hydrographic observations. *Scientific Data*, *13*(1), 456. https://doi.org/10.1038/s41597-026-06749-8

Olsen, A., Lange, N., Key, R. M., Tanhua, T., Bittig, H. C., Kozyr, A., et al. (2020). An updated version of the global interior ocean biogeochemical data product, GLODAPv2.2020. *Earth System Science Data*, *12*(4), 3653–3678. https://doi.org/10.5194/essd-12-3653-2020

Portela, E., Kolodziejczyk, N., Maes, C., & Thierry, V. (2020). Interior water-mass variability in the Southern hemisphere Oceans during the last decade. *Journal of Physical Oceanography*, *50*(2), 361–381. https://doi.org/10.1175/JPO-D-19-0128.1

Pritchard, H. D., Ligtenberg, S. R., Fricker, H. A., Vaughan, D. G., Van Den Broeke, M. R., & Padman, L. (2012). Antarctic ice-sheet loss driven by basal melting of ice shelves. *Nature*, *484*(7395), 502–505. https://doi.org/10.1038/nature10968

Purkey, S. G., & Johnson, G. C. (2010). Warming of global abyssal and deep Southern Ocean waters between the 1990s and 2000s: Contributions to global heat and Sea level rise budgets. *Journal of Climate*, *23*(23), 6336–6351. https://doi.org/10.1175/2010JCLI3682.1

Purkey, S. G., & Johnson, G. C. (2012). Global contraction of Antarctic bottom water between the 1980s and 2000s. *Journal of Climate*, *25*(17), 5830–5844. https://doi.org/10.1175/JCLI-D-11-00612.1

Qu, T., Gao, S., & Fine, R. A. (2020). Variability of the sub–Antarctic mode water subduction rate during the Argo period. *Geophysical Research Letters*, *47*(13), e2020GL088248. https://doi.org/10.1029/2020GL088248

Rahmstorf, S., Box, J. E., Feulner, G., Mann, M. E., Robinson, A., Rutherford, S., & Schaffernicht, E. J. (2015). Exceptional twentieth-century slowdown in Atlantic Ocean overturning circulation. *Nature Climate Change*, *5*(5), 475–480. https://doi.org/10.1038/nclimate2554

Rhein, M., Kieke, D., & Steinfeldt, R. (2015). Advection of north Atlantic deep water from the Labrador Sea to the southern hemisphere. *Journal of Geophysical Research: Oceans*, *120*(4), 2471–2487. https://doi.org/10.1002/2014JC010605

Roemmich, D., Johnson, G., Riser, S., Davis, R., Gilson, J., Owens, W. B., et al. (2009). The Argo Program: Observing the global Oceans with profiling floats. *Oceanography*, *22*(2), 34–43. https://doi.org/10.5670/oceanog.2009.36

Romera-Castillo, C., Álvarez, M., Pelegrí, J. L., Hansell, D. A., & Álvarez-Salgado, X. A. (2019). Net additions of recalcitrant dissolved organic carbon in the deep Atlantic Ocean. *Global Biogeochemical Cycles*, *33*(9), 1162–1173. https://doi.org/10.1029/2018GB006162

Romera-Castillo, C., Heras, J., Álvarez, M., Álvarez Salgado, X. A., Mata, G., & Sáenz-de Cabezón, E. (2022). Application of multi-regression machine learning algorithms to solve ocean water mass mixing in the Atlantic Ocean. *Frontiers in Marine Science*, *9*, 904492. https://doi.org/10.3389/fmars.2022.904492

Sandoval, F. J., & Weatherly, G. L. (2001). Evolution of the deep Western boundary Current of Antarctic bottom water in the Brazil Basin. *Journal of Physical Oceanography*, *31*(6), 1440–1460. https://doi.org/10.1175/1520-0485(2001)031⟨1440:EOTDWB⟩2.0.CO;2

Sarmiento, J. L., Gruber, N., Brzezinski, M. A., & Dunne, J. P. (2004). High-latitude controls of thermocline nutrients and low latitude biological productivity. *Nature*, *427*(6969), 56–60. https://doi.org/10.1038/nature02127

Schmidtko, S., Heywood, K. J., Thompson, A. F., & Aoki, S. (2014). Multidecadal warming of Antarctic waters. *Science*, *346*(6214), 1227–1231. https://doi.org/10.1126/science.1256117

Schodlok, M. P., Menemenlis, D., & Rignot, E. J. (2016). Ice shelf basal melt rates around Antarctica from simulations and observations. *Journal of Geophysical Research: Oceans*, *121*(2), 1085–1109. https://doi.org/10.1002/2015JC011117

Shrikumar, A., Lawrence, R., & Casciotti, K. L. (2022). PYOMPA technical note. https://doi.org/10.1002/essoar.10507053.4

Silvano, A., Purkey, S., Gordon, A. L., Castagno, P., Stewart, A. L., Rintoul, S. R., et al. (2023). Observing Antarctic bottom water in the Southern Ocean. *Frontiers in Marine Science*, *10*, 1221701. https://doi.org/10.3389/fmars.2023.1221701

Sloyan, B. M., & Rintoul, S. R. (2001). Circulation, renewal, and modification of Antarctic mode and intermediate water. *Journal of Physical Oceanography*, *31*(4), 1005–1030. https://doi.org/10.1175/1520-0485(2001)031⟨1005:CRAMOA⟩2.0.CO;2

Sloyan, B. M., Wanninkhof, R., Kramp, M., Johnson, G. C., Talley, L. D., Tanhua, T., et al. (2019). The Global Ocean Ship-Based Hydrographic Investigations Program (GO-SHIP): A platform for integrated multidisciplinary Ocean science. *Frontiers in Marine Science*, *6*, 445. https://doi.org/10.3389/fmars.2019.00445

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

Talley, L. (2013). Closure of the global overturning circulation through the Indian, Pacific, and Southern Oceans: Schematics and transports. *Oceanography*, *26*(1), 80–97. https://doi.org/10.5670/oceanog.2013.07

Talley, L. D. (1996). Antarctic intermediate water in the south Atlantic. In *The south Atlantic* (pp. 219–238). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-80353-6_11

Talley, L. D. (2011). *Descriptive physical oceanography: An introduction* (6th ed. ed.). Elsevier Science and Technology.

Tamsitt, V., England, M. H., Rintoul, S. R., & Morrison, A. K. (2021). Residence time and transformation of warm circumpolar deep water on the Antarctic Continental Shelf. *Geophysical Research Letters*, *48*(20), e2021GL096092. https://doi.org/10.1029/2021GL096092

Thomas, J. L., Waugh, D. W., & Gnanadesikan, A. (2015). Southern Hemisphere extratropical circulation: Recent trends and natural variability. *Geophysical Research Letters*, *42*(13), 5508–5515. https://doi.org/10.1002/2015GL064521

Thompson, R. O. R. Y., & Edwards, R. J. (1981). Mixing and water-mass Formation in the Australian subantarctic. *Journal of Physical Oceanography*, *11*(10), 1399–1406. https://doi.org/10.1175/1520-0485(1981)011⟨1399:MAWMFI⟩2.0.CO;2

Tomczak, M. (1981). A multi-parameter extension of temperature/salinity diagram techniques for the analysis of non-isopycnal mixing. *Progress in Oceanography*, *10*(3), 147–171. https://doi.org/10.1016/0079-6611(81)90010-0

Toole, J., Curry, R., Joyce, T., McCartney, M., & Peña-Molino, B. (2011). Transport of the North Atlantic Deep Western Boundary Current about 39°N, 70°W: 2004–2008. *Deep Sea Research Part II: Topical Studies in Oceanography*, *58*(17–18), 1768–1780. https://doi.org/10.1016/j.dsr2.2010.10.058

Yashayaev, I. (2024). Intensification and shutdown of deep convection in the Labrador Sea were caused by changes in atmospheric and freshwater dynamics. *Communications Earth and Environment*, *5*(1), 156. https://doi.org/10.1038/s43247-024-01296-9

Yashayaev, I., & Loder, J. W. (2016). Recurrent replenishment of Labrador Sea Water and associated decadal-scale variability: 2015 Convection in Labrador Sea. *Journal of Geophysical Research: Oceans*, *121*(11), 8095–8114. https://doi.org/10.1002/2016JC012046