



The Spatiotemporal Distribution of Dissolved Inorganic Carbon in the Global Ocean Interior: Reconstructed through Machine Learning

Tobias Ehmen^{1,2}, Neill Mackay², and Andrew J Watson²

¹University of Bergen and Bjerknes Centre for Climate Research, Bergen, Norway

²University of Exeter, Exeter, United Kingdom

Correspondence: Tobias Ehmen (tobias.ehmen@uib.no)

Abstract.

The oceans mitigate climate change by absorbing 25-30% of anthropogenic carbon emissions. Decadal variability in the ocean carbon sink has been suggested by pCO₂-based reconstructions of the ocean uptake, but these variations are less apparent in reconstructions based on global ocean biogeochemistry models (GOBMs) which also predict a slower increase in the ocean sink over the period from 2000 to 2015, raising concerns about our ability to accurately project future changes. An independent method to study the uptake uses interior observations to calculate how dissolved inorganic carbon (DIC) is changing. To address the sparsity of interior data, machine learning techniques have been applied, but as yet, global, full-depth reconstructions of the complete inventory of DIC have not been produced. Here we develop such a full-depth reconstruction, from the 1990s to 2019, using "ResNet-DIC", a deep neural network trained on GLODAPv2.2023 observations. Atmospheric CO₂, location, temperature, and salinity from EN4 analysis are used as predictors. Our method shows good predictive power when validated using independent data sets and model subsampling experiments. We diagnose the total oceanic DIC pool as $37,508 \pm 274$ Pg C in 2019 with an average rate of increase of 2.64 ± 0.13 Pg C/year over the whole time period. The global change in the DIC inventory exhibits pronounced peaks in decadal variability, especially in the early 2000s, driven primarily by intermediate waters at depths of 300-1000 m, particularly in the Atlantic, Indian, and Southern Oceans, and to a lesser extent in the Pacific. The accumulation rate of DIC increases steadily from the mid-2000s.

1 Introduction

The oceans absorb 25-30% of the anthropogenic carbon that is released into the atmosphere by human activities (Friedlingstein et al., 2026; Gruber et al., 2023), thus mitigating climate change (Intergovernmental Panel on Climate Change, 2023; Friedlingstein et al., 2026). This uptake results in changes to the ocean chemistry including ocean acidification (Orr et al., 2005; Kwiatkowski et al., 2020), affecting marine organisms and ecosystems (Orr et al., 2005; Fabry et al., 2008; Doney et al., 2009). Müller and Gruber (2024) highlighted that acidification extends to the whole interior ocean and that ecosystems of deeper ocean layers may be even more impacted by this process than organisms at the surface (Seibel and Walsh, 2001; Hennige et al., 2020).



Carbon in the ocean is mainly stored as dissolved inorganic carbon (DIC), representing carbon in the forms of bicarbonate, carbonate and dissolved CO₂. The total DIC pool in the ocean holds ~90% of the total carbon budget of the earth's biosphere and is currently estimated as ~37,000 PgC (Keppler et al., 2020). DIC is frequently divided into C_{nat}, a natural, and C_{ant}, an anthropogenic component. C_{nat} represents the carbon pool that existed prior to the introduction of anthropogenic CO₂, where for an assumed pre-industrial global steady state, an outgassing flux at the air-sea interface was balanced by an input of terrestrial carbon transported through rivers (Regnier et al., 2022; Gruber et al., 2023). This natural steady state also reflects the action of the biological carbon pump, which transports carbon from the surface to the deep ocean and plays a central role in setting the vertical distribution of C_{nat}. Anthropogenic carbon, C_{ant}, perturbs this steady state, as accumulating anthropogenic CO₂ in the atmosphere enters the ocean through air-sea exchange (Sarmiento and Sundquist, 1992; Friedlingstein et al., 2026; Gruber et al., 2023). This means that the increase in DIC can be seen as a proxy for the ocean's uptake of anthropogenic CO₂, at least on multidecadal timescales, provided that changes in the natural carbon cycle - particularly the strength of the biological carbon pump - are small relative to the anthropogenic signal (Sarmiento and Gruber, 2013; Gruber et al., 2023).

Ocean-atmosphere carbon fluxes are increasingly thought to be more variable than previously supposed (Gruber et al., 2023). This is especially the case with further increases in the atmospheric CO₂ concentration, as the uptake of carbon is strongly related to the growth rate of atmospheric CO₂ (McKinley et al., 2020). The spatiotemporal evolution of the ocean's carbon budget is not only difficult to assess because of current limitations in models (Verdy and Mazloff, 2017; Terhaar et al., 2024) but also because of the sparseness of observations (Hauck et al., 2023). Quality controlled biogeochemical datasets documenting ocean carbon, such as SOCAT for the surface ocean pCO₂ (Bakker et al., 2016) and GLODAP for the interior ocean (Lauvset et al., 2024) represent most available measurements. However, all available pCO₂ datapoints in the SOCAT database, for example, cover merely 2% of grid cells on a global monthly 1°x1° grid from 1990 (Hauck et al., 2023). In both datasets, the data are unevenly distributed in space and time, with fewer high-quality datapoints prior to the mid-1980s (Wallace, 1995), and they also have a strong seasonal bias, with coverage especially sparse in polar regions and during the winter months (Bakker et al., 2016; Lauvset et al., 2016).

To overcome the challenge of sparse observations, a large part of interpolation and reconstruction efforts is now driven by machine learning methods (Landschützer et al., 2013, 2020; Laruelle et al., 2017; Keppler et al., 2020, 2023; Zemskova et al., 2022; Sharp et al., 2023; Ford et al., 2024; Ito et al., 2024), which cope well with sparse, non-linear data like ocean biogeochemistry observations. Previous reconstructions of the air-sea CO₂ flux, which represents the main pathway of adding anthropogenic CO₂ into the DIC pool, are based on pCO₂ measurements (Landschützer et al., 2013, 2016, 2020; Fay et al., 2021). They have suggested that there is a decadal variability in the ocean's uptake of carbon, with an apparent weakening of the ocean's carbon sink in the 1990s (Le Quéré et al., 2007) and a strengthening back to the levels expected for the long-term uptake in the 2000s (Landschützer et al., 2014, 2020; Gruber et al., 2023). The reasons for such variability are not well understood (Landschützer et al., 2016) and there remains a possibility that it is an artifact of the sparse biogeochemical data on which the pCO₂ products and other analyses are based (Gloege et al., 2021). If the variability it is real, it might be explained by varying atmospheric CO₂ increase (McKinley et al., 2020), atmospheric circulation changes (Le Quéré et al., 2007; Landschützer et al., 2015; Keppler and Landschützer, 2019) or changes in the overturning circulation of the upper ocean (DeVries and Weber,



2017). Moreover, the magnitude of the variability and trend suggested by the $p\text{CO}_2$ observations is currently not represented in GOBMs (DeVries et al., 2019; Hauck et al., 2020), and mean model-based estimates of the ocean accumulation of C_{ant} are consistently lower than those from observation-based products, underestimating the carbon sink by 10-20% (Hauck et al., 2020; Friedlingstein et al., 2026).

Reconstructions of the interior ocean carbon, using interior observations of the ocean, which are also sparse, are an alternative method to investigate the variability of the ocean's carbon sink. Previous reconstructions have either been restricted to the upper ocean or the Southern Ocean (Keppler et al., 2020, 2023; Zemskova et al., 2022), although TRACEv1 recently provided a global, full-depth reconstruction of anthropogenic carbon in the ocean interior (Carter et al., 2025). A global, full-depth reconstruction of DIC derived directly from quality-controlled interior carbon observations is still lacking. Furthermore, previous products have not rigorously used quality controlled biogeochemical samples in the training process. Instead, they either used smoothed climatologies (Keppler et al., 2020, 2023) or various surface parameters of remote sensing data from the Southern Ocean for training (Zemskova et al., 2022).

Here, we present a new reconstruction of the evolution of the ocean's DIC inventory at nearly global coverage and to 5500 m depth, covering the period 1990-2019. The full-depth variation of the DIC field contains valuable information about the ocean's ability to take up and store carbon, including C_{ant} (Gruber and Sarmiento, 2002; Friedlingstein et al., 2026). It is also an important measure of the magnitude and distribution of interior ocean acidification (Gruber et al., 2023). Furthermore, knowledge of how the DIC distribution in the interior ocean changed over time helps understand the mechanisms that facilitate the transport of carbon in the ocean. Transport from the ocean's surface to deeper layers is an important factor in modulating the uptake of anthropogenic CO_2 and is key to reduce uncertainties in our estimates of the future evolution of the ocean's carbon sink in response to climate change (Sarmiento et al., 1992; Matear and Hirst, 1999; Intergovernmental Panel on Climate Change, 2023; Gruber et al., 2023).

We use a residual neural network, a type not previously used for ocean carbon studies. We use primarily temperature and salinity as predictor variables, which greatly increases the availability of observations used for prediction. We then validate our results using independent time series and synthetic data from two ocean biogeochemical models. The reconstruction is designed to capture interannual and decadal variability in the ocean interior carbon inventory, rather than seasonal fluctuations, in order to (1) assess to what extent the global carbon sink variability suggested by surface $p\text{CO}_2$ products agrees with estimates from the interior ocean and (2) identify the regions driving interior ocean carbon inventory changes as a component of the global carbon budget.

2 Data and Methods

2.1 ResNet-DIC mapping

Here, we apply a ResNet-based machine learning model, adapted for dissolved inorganic carbon and referred to as ResNet-DIC, to reconstruct DIC concentrations in four dimensions (location, depth, and time). This architecture has also recently been used in related oceanographic applications, including Antarctic salinity reconstruction (Sohail et al., 2026). Neural networks



can identify complex patterns unconstrained by assumptions of linearity, and are thus ideal for modelling non-linear systems, while deep networks can find more complex patterns than shallow networks (He et al., 2015). ResNet-DIC contrasts with the SOM-FFM approach (Landschützer et al., 2013) that has been applied to interpolate pCO₂-based air-sea CO₂ flux estimates (Landschützer et al., 2020; Ford et al., 2024) and to create the MOBO-DIC product, an interior ocean DIC reconstruction (Keppler et al., 2020, 2023). The SOM-FFN method uses self-organising maps, SOM (Kohonen, 1989, 2001), to divide all data into biogeochemical biomes prior to use of a shallow feed forward neural network (FFN) that finds appropriate solutions for each biome. These are then combined into a single reconstruction. A deeper neural network with many levels can be trained directly, without the need to separate and average the data from different pre-defined biomes. However, increasing the number of hidden layers in a feed-forward neural network can make optimization more difficult and hinder convergence. We therefore employ a Residual Neural Network (ResNet) architecture.

A ResNet is a type of deep neural network that is designed to capture both low-level and high-level features effectively (He et al., 2015). ResNets are characterised by skip connections that add the unchanged information of previous hidden layers to the information that flows through additional hidden layers. This design makes it easier for the neural network to converge, making it feasible to train deeper architectures. Our ResNet architecture uses three groups of layers. Each of those groups has three hidden layers with one skip connection (see figure 1a). This architecture improved the prediction of the validation dataset (for more details on the validation dataset see section 2.3), reducing errors by 15 to 30% compared to shallower neural networks, with average RMSE during early testing dropping from ~12 µmol/kg using a "vanilla" neural network to ~9 µmol/kg using ResNet. The size of each hidden layer was determined using the hyperparameter testing function in Tensorflow (TensorFlow Developers, 2024), where neural network architectures can be tested systematically.

ResNet-DIC is trained on a recent version of the Global Data Analysis Project dataset (GLODAPv2.2023), a quality-controlled collection of biogeochemical data (Lauvset et al., 2024). After training, the prediction is then mapped on the temperature and salinity fields of the EN4 analysis product of the UK Met Office, which provides objective analysis of temperature and salinity on a regular time and spatial grid (Good et al., 2013). The resulting product has the same shape, resolution, and range as the EN4 analysis grid: a monthly near-global 1°x1° regular grid from -83° to 89° latitude with the deepest depth cells reaching 5500 m. The approach of mapping on EN4 is similar to the observation-based product for oxygen by Ito et al. (2024). The EN4 analysis is available from 1900 to 2026, with varying uncertainties. We choose to predict from the year 1990 onwards, due to the increasing number of high quality temperature and salinity profiles that are available from that time.

Several training control mechanisms are implemented to improve learning and reduce overfitting. A subset of the data is reserved for validation during training (see section 2.3). If validation error plateaus, the learning rate is reduced to allow further optimisation using smaller parameter updates. Training is terminated through early stopping once no further improvement in validation error is observed, which prevents overfitting. Final reconstructions are generated as the average of 15 ensemble members, each trained using a differently randomised data partitioning. This reduces sensitivity to any individual validation split and additionally enables uncertainty estimation (section 3.3). Prior to training we use z-score normalisation, which normalises each variable to a mean value of 0 and standard deviation of 1. This ensures that the input data have a consistent scale, facilitating faster and more stable convergence of the neural network. Model performance with this approach was strong, indicating

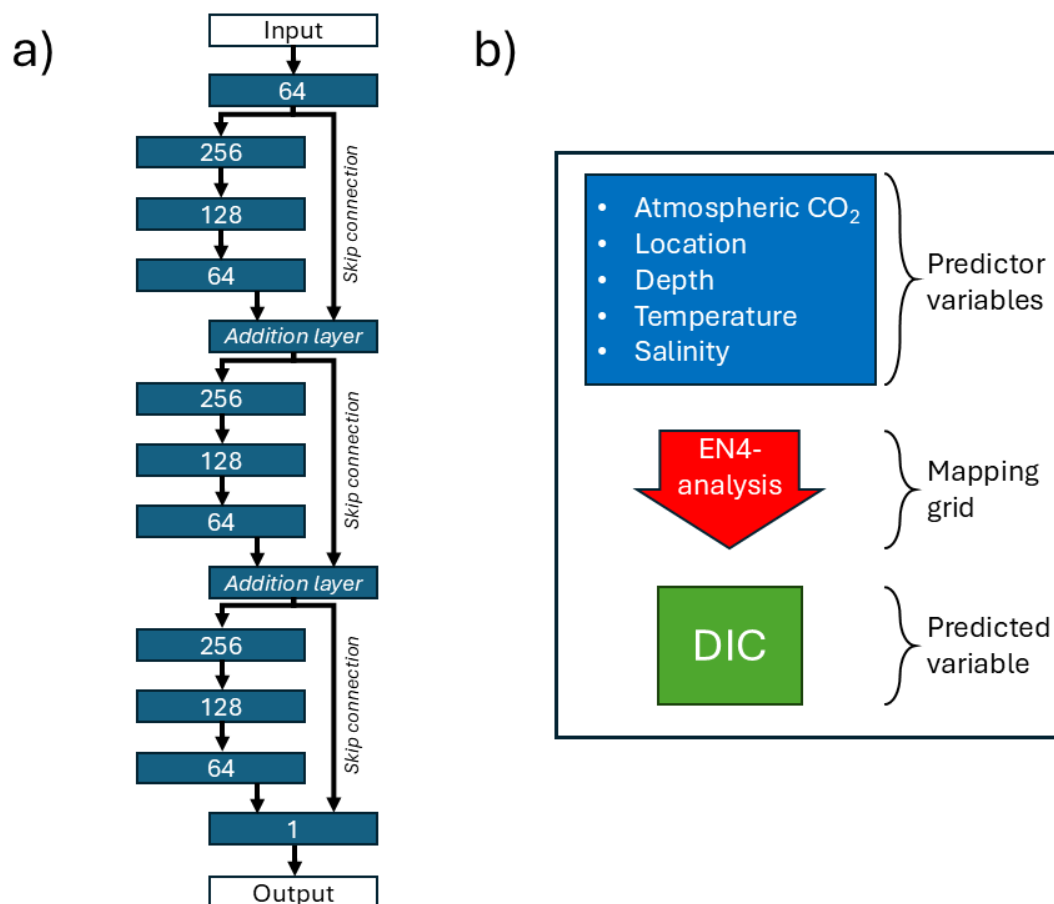


Figure 1. a) Model architecture of ResNet model used in DIC prediction. Numbers show the neurons in each layer. Note the skip connections that add the information of previous hidden layers to the information that flowed through additional hidden layers; b) schematic of input and output variables and mapping grid.

that the presence of a few instances of skewness or outliers did not adversely impact results, and thus z-score normalization remained an appropriate choice.

2.2 Predictor variables

130 In choosing the variables with which to train the network we considered the trade-off between relying principally on relatively abundant temperature and salinity observations that accompany the great majority of DIC data, or additionally incorporating nutrients and oxygen, which might improve predictions, but would greatly reduce the instances of collocated, concurrent observations. The MOBO-DIC product (Keppler et al., 2020, 2023) used temperature, salinity, dissolved oxygen, nitrate, and silicic acid as predictors. However, monthly climatologies from the World Ocean Atlas 2018 (Boyer et al., 2018) were used in
 135 both training and reconstruction, whereas measured DIC values from GLODAPv2.2021 (Lauvset et al., 2021) were used only



during training as the target variable for model prediction. The use of climatologies decreased the noise in the reconstructions and avoided reducing co-measured data to only a fraction of the GLODAP dataset, as very few locations have all predictors simultaneously available. However, training a neural network on interpolated climatologies risks overly smooth reconstructions, may incorporate errors and biases present in the climatologies, and may cause variability to be underestimated.

140 Instead, we use temperature and salinity as our primary predictors for ResNet-DIC, in addition to variables representing time and space (see figure 1b). Temperature and salinity have a strong impact on the carbon distribution in the ocean. While temperature largely controls CO₂ solubility (Weiss, 1974), salinity impacts the ocean's alkalinity, which regulates the buffering capacity of seawater and thus the storage of DIC (Williams et al., 2011). Temperature, salinity, and pressure are crucial parameters governing ocean circulation and stratification, and because temperature and salinity act as conservative tracers of
 145 circulation and mixing, they are well suited for reconstructing DIC (Turner et al., 2023).

Other predictors we use are atmospheric CO₂, location, and depth. We take monthly mean atmospheric CO₂ concentrations from Mauna Loa (deseasonalised), serving both as a proxy for the temporal dimension - effectively replacing the need for an explicit time variable such as decimal year - and as a key driver of air-sea CO₂ exchange. We found that using atmospheric CO₂ in this way provides better generalisation than directly using time. For location, we use an n-vector (Gade, 2010), a
 150 three component vector pointing away from the centre of an ellipsoid earth, that is unique to location and resolves the spatial discontinuity at 0 and 360 degrees longitude. This also bypasses issues noted by Keppler et al. (2023) with using cosine and sine curves for longitude and latitude.

By predicting primarily using temperature and salinity, we limit the use of climatologies based on very sparse data during the mapping process. The predictions are mapped on the EN4 analysis product for temperature and salinity. Acknowledging
 155 the caveat that EN4 relaxes to a climatology where there are no observations available (Good et al., 2013), we have made this choice in preference to directly training the model on climatological fields or monthly mapped temperature and salinity products derived from Argo float observations, as done for MOBO-DIC, to achieve our aim of a global, full-depth reconstruction. Keppler et al. (2023) found that using co-measured discrete predictors resulted in noisier reconstructions and therefore adopted mapped predictor fields. Since we are focused on interannual or decadal variability, rather than seasonal variability, we consider
 160 the noisiness associated with training on discrete observations not to present a problem. Our priority is for the training process itself to remain constrained by discrete observational data rather than by spatially smoothed climatological estimates.

The comparatively good data availability of temperature and salinity, with O(10⁶) available datapoints in the EN4 profiles dataset (Good et al., 2013), increases confidence in the gridded temperature and salinity values compared to other parameters like oxygen or nutrients. Oxygen, one of the most sampled biogeochemical parameters, has approximately O(10⁵) datapoints
 165 available (Garcia et al., 2019; Grégoire et al., 2021), an order of magnitude fewer. Although the EN4 analysis extends back to 1900, we restrict our analysis to periods from the 1990s onwards to avoid times when the product is predominantly based on climatology due to sparse observations (see figure A1), thereby ensuring a minimum level of data coverage.

If we forego the computationally-intensive reconstruction phase, we can train ResNet-DIC using any combination of predictor variables available in the training and validation datasets. Figure 2 shows the results of the validation process for different
 170 combinations of predictor variables, with root mean squared error (RMSE) and bias reported as performance metrics. The

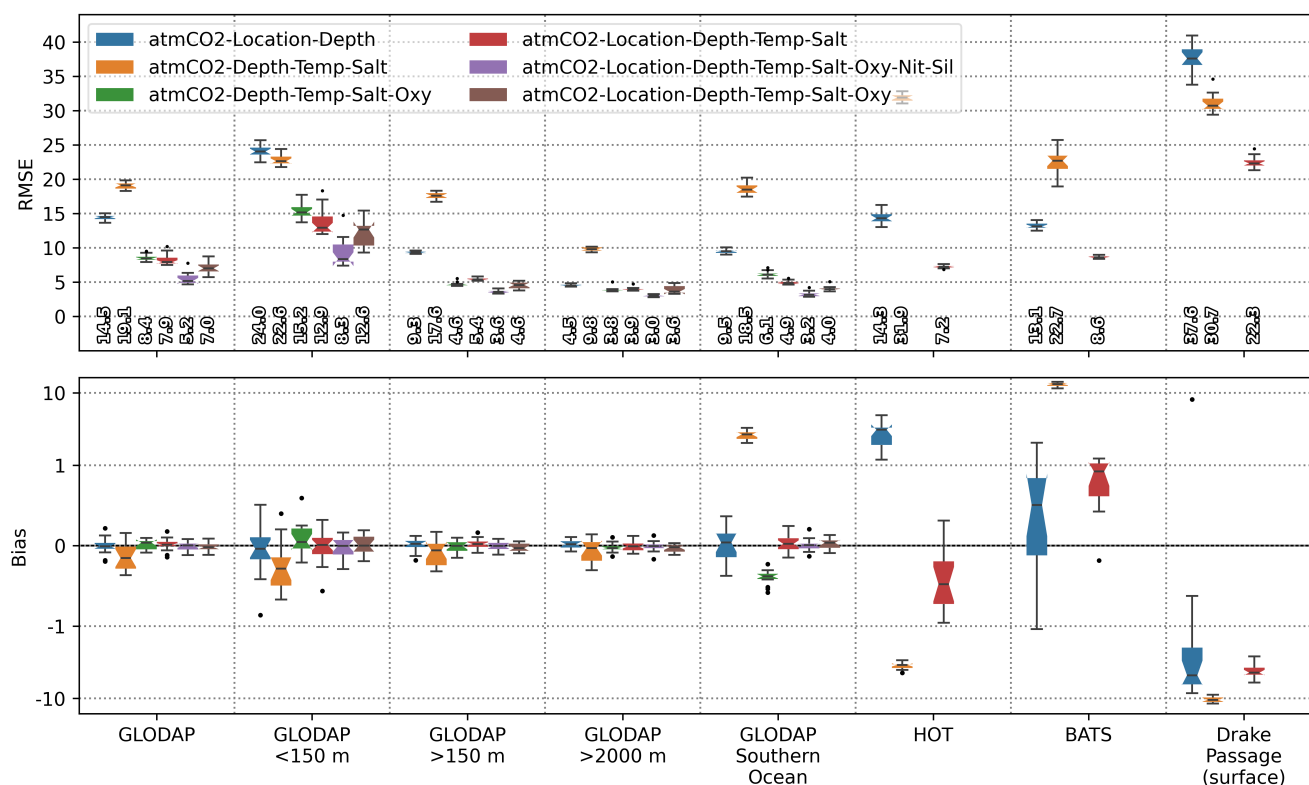


Figure 2. Boxplot comparison of model prediction errors using different combinations of predictor variables: atmospheric CO₂, Location, Depth, Temperature, and Salinity (red – the configuration used for the final DIC product); atmospheric CO₂, Location, Depth (blue); atmospheric CO₂, Depth, Temperature, and Salinity (orange); atmospheric CO₂, Depth, Temperature, Salinity, and Oxygen (green); atmospheric CO₂, Location, Depth, Temperature, Salinity, and Oxygen (brown); and atmospheric CO₂, Location, Depth, Temperature, Salinity, Oxygen, Silicate and Nitrate (purple). Boxes represent the interquartile range (25th–75th percentile), with the median shown as a horizontal line and whiskers extending to 1.5 times the interquartile range. Omitting either Location or Temperature and Salinity leads to increased errors and biases, indicating their importance as predictors. Adding Oxygen marginally reduces errors and can also be a substitute for location.

estimates including nutrients and oxygen use smaller sample sizes than those using only temperature and salinity, but that should not affect these comparisons. The predictor set we finally chose, with atmospheric CO₂, location, depth, temperature, and salinity, resulted in low overall errors, with mean RMSEs of $7.9 \pm 0.8 \mu\text{mol/kg}$ compared to GLODAP. When temperature and salinity were excluded, errors increased substantially across all datasets - for example, GLODAP RMSE rose to $14.5 \pm 0.4 \mu\text{mol/kg}$, HOT increased to $14.3 \pm 0.9 \mu\text{mol/kg}$ and BATS to $13.1 \pm 0.5 \mu\text{mol/kg}$ (see figure 2).

Bias patterns largely followed the RMSE trends, with the lowest biases observed using the target predictor set and notable increases when either temperature and salinity or location were omitted. Adding oxygen or silicate and nitrate had only minor effects on bias, consistent with its modest impact on overall error.



Adding oxygen as an additional predictor led to modest improvements, reducing the overall GLODAP RMSE from 7.9 ± 0.8 to $7.0 \pm 0.8 \mu\text{mol/kg}$, with similar trends across the Southern Ocean and other validation datasets. Including silicate and nitrate, as in Keppler et al. (2020), further improved performance, lowering the overall GLODAP RMSE to $5.2 \pm 0.7 \mu\text{mol/kg}$. While these additions are clearly beneficial for mapping onto global grids this must be weighed against the sparser availability and additional uncertainty associated with oxygen data. This uncertainty increases further if also adding silicate and nitrate.

Excluding location information degraded model performance across all validation datasets (figure 2). For GLODAP, RMSE increased from 7.9 ± 0.8 to $19.1 \pm 0.5 \mu\text{mol/kg}$, and errors below 2000 m more than doubled compared to the other predictor combinations. This indicates that spatial context is critical for accurate predictions in these datasets, especially to distinguish between water masses that are geographically separated but share similar temperature and salinity characteristics. A related but distinct behavior is observed in Keppler et al. (2023), where including location did not improve performance when oxygen, nitrate, and silicate were already included as predictors. Rather than being contradictory, this likely reflects differences in how spatial information is represented: in MOBO-DIC, these biogeochemical predictors are derived from mapped climatological fields and therefore already contain large-scale spatial structure, reducing the added value of explicit location coordinates. We provide further evidence for this by adding oxygen as a predictor while excluding location. The GLODAP RMSE decreased again to $8.4 \pm 0.5 \mu\text{mol/kg}$, close to the values obtained using location, temperature, and salinity. This suggests that oxygen - likely also when further adding nitrate and silicate - can partly compensate for the absence of explicit location predictors, making this interpretation consistent with MOBO-DIC. These comparisons emphasise that both location and physical tracers such as temperature and salinity play key roles in model performance, especially if a prediction with oxygen and nutrients through the use of climatologies is to be avoided. Further information on predictor importance is provided in section A1.

2.3 Training dataset, partitioning, and validation strategy

Training is restricted to GLODAPv2.2023 observations flagged as “good” according to both the original WOCE quality control procedures (quality flag of 2) and the GLODAP secondary quality control procedures (quality flag of 1). Where bathymetry permits, GLODAP provides substantial full-depth coverage, and represents a dataset focused on the interior ocean. For DIC measurements where all flags are applied, 26.3% of datapoints are found in depths greater than 2000 m (100,927 points), while 73.7% are from shallower depths (283,140 points), demonstrating that the deep ocean is meaningfully sampled in the training data.

We apply a 15% train/test split, where complete vertical profiles are randomly selected and held out as a validation dataset. Splitting is performed at the vertical profile level rather than the individual measurement level to ensure a more realistic generalisation challenge, as withholding isolated points within profiles would allow trivial reconstruction through vertical interpolation. To further assess model generalisation, we additionally retain a fully independent 10% hold-out test dataset that is never used during training or early stopping.

A key concern in neural network training is potential overfitting to the training or validation data. We therefore explicitly analyse training dynamics using error loss curves across the training, validation, and independent hold-out test datasets (Figure 3). These curves show no evidence of overfitting: validation and hold-out losses closely track each other throughout training,

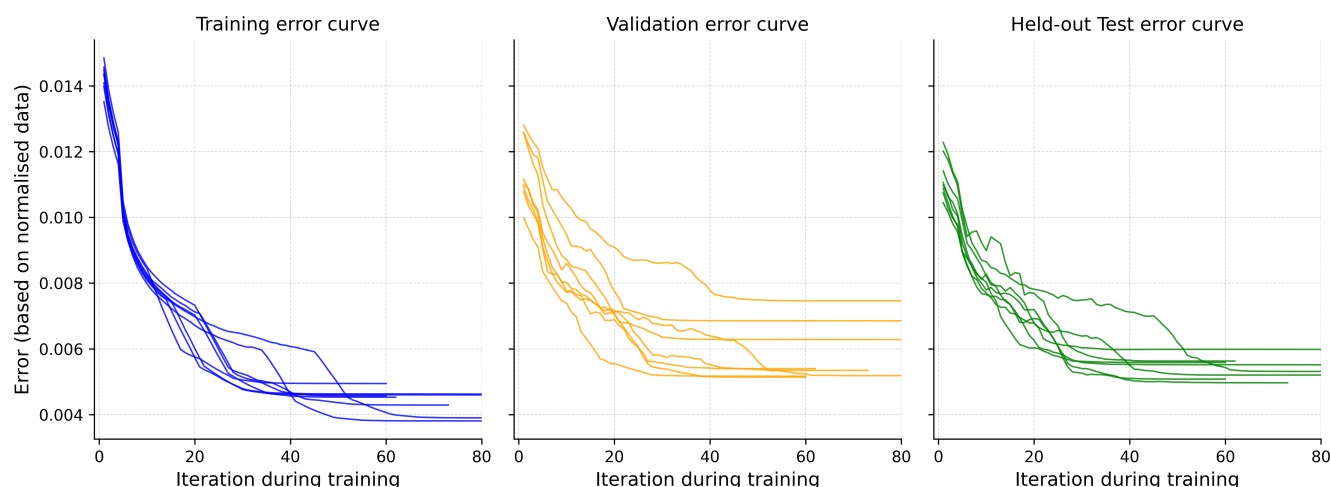


Figure 3. Error loss curves during training for a ensemble trained on GLODAP data. Curves are displayed for the training dataset (left panel), which is used for training; the validation dataset (middle panel), consisting of 15% of data that is seen during training to guide learning and prevent overfitting; and the test dataset (right panel), which is never seen in training. Note that as the data and error are scaled to normalised values.

follow nearly identical convergence trajectories across ensemble members, and stabilise at comparable final loss values. While small stochastic fluctuations occur during intermediate epochs, no divergence between datasets is observed, indicating stable
 215 generalisation behaviour.

We also find that training dynamics are highly consistent across validation and hold-out test data, as shown in the error loss curves (figure 3). Both datasets follow very similar trajectories across ensemble members, indicating that the validation set is representative and that no systematic overfitting to the validation data occurs. Based on this behaviour, we use the 15% validation set for model selection, performance reporting and final predictions, as it provides a stable and representative
 220 estimate of out-of-sample error while preserving maximum training data availability.

The overall fit between the prediction and the measured values in the 15% validation data of each run is encouraging (see figure 4). For the GLODAP validation dataset through splitting by complete vertical profiles, ResNet-DIC achieves an RMSE of $9.8 \pm 3.1 \mu\text{mol/kg}$ on average and a bias of $-0.07 \pm 0.2 \mu\text{mol/kg}$ across all ensemble members (see figure A2). Values of the coefficient of determination, R^2 , often exceed 0.98, indicating an excellent fit between predicted and observed values.
 225 Compared over the same depth range as Keppler et al. (2023), where MOBO-DIC achieved an RMSE of $16 \mu\text{mol/kg}$ with a bias of $0 \mu\text{mol/kg}$ when predicting the GLODAP dataset, ResNet-DIC results in an RMSE of $11.5 \pm 3.7 \mu\text{mol/kg}$ and a bias of $-0.10 \pm 0.3 \mu\text{mol/kg}$ when split along complete vertical profiles.

Differences between profile- and cruise-based splitting emerge when evaluating held-out validation data from the observational dataset (Figure A2), with cruise-based splitting leading to reduced predictive performance. Cruise-based splitting
 230 provides a more stringent test of generalisation by withholding entire sampling campaigns, thereby increasing the geographic

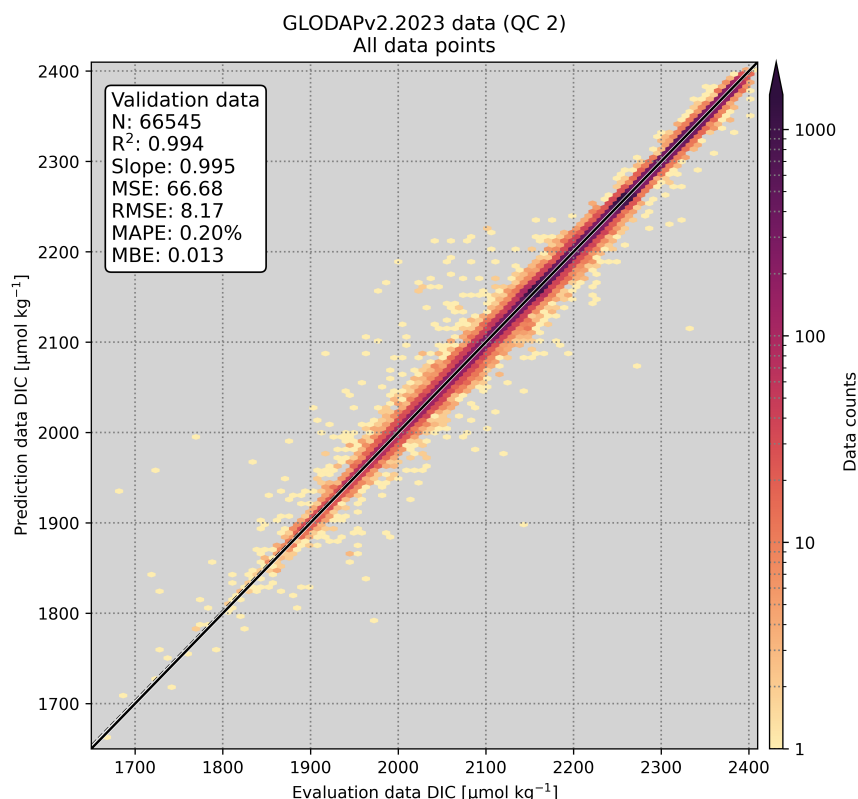


Figure 4. Comparison plot of predicted DIC against true DIC from the GLODAP validation dataset (single ensemble member). This validation dataset results from withholding 15% of the training data, splitting along vertical profiles.

and environmental separation between training and validation data. Because the primary objective is global reconstruction rather than prediction for entirely unseen campaigns, we adopt the vertical-profile split as the standard configuration.

3 Validation and uncertainty estimation

3.1 Validation using model output

235 Independent datasets based on observations have the disadvantage that they do not cover the global ocean evenly and consistently. Where no data are present, there is a high chance that errors and biases in our predictions are not revealed by the validation we have outlined so far. To test how our model predictions might perform in a fully sampled global ocean - and to evaluate how representative the observational training dataset is - we conduct a sampling experiment using the ECCO-Darwin and NorESM models. ECCO-Darwin is a global ocean biogeochemistry state estimate that assimilates physical and biogeochemical observations. It conserves physical and biogeochemical tracers and reproduces global CO_2 flux variability in line

240



with observation-based products (Carroll et al., 2020). NorESM is an Earth system model with an interactive ocean biogeochemistry component, providing a fully coupled representation of climate–carbon cycle interactions and internally generated ocean biogeochemical variability (Seland et al., 2020). Because the historical NorESM simulation ends in 2014 we use the SSP5-8.5 scenario simulation to maintain coverage over the full observational period. While neither of them is a perfect representation of the real ocean, these models provides a physically consistent, fully gridded reference field against which the effects of observational sampling can be tested.

Model-based validation is performed by constructing a synthetic dataset: gridded monthly output of ECCO-Darwin and NorESM is subsampled at the locations and times of GLODAP observations, creating a sparse dataset analogous to the real-world training data. Monthly averages from ECCO-Darwin are matched to the nearest GLODAP timestamps and positions. To preserve consistency with the model “truth,” no interpolation is performed in depth; instead, the nearest available model value is selected. Duplicates are retained to preserve the data density and distribution of the original GLODAP dataset, even if they do not provide new information. Most GLODAP observations are represented in the synthetic datasets. The NorESM historical and SSP5-8.5 simulations span the full observational period, whereas the ECCO-Darwin-based synthetic dataset omits only a small proportion of observations outside its 1985–2021 time window, resulting in closely matched data densities to the observational dataset overall (figure A3).

The synthetic dataset is then used to train separate neural networks, following the same approach as for the real-world data. Following training, the full gridded fields of the models are predicted. Because the true (i.e., fully sampled) model fields are known, this setup allows us to evaluate how well a sparsely sampled dataset - such as GLODAP - is suited to reproducing global-scale variability. Differences between the synthetic and full-model predictions offer valuable insights into where and to what extent observational coverage may limit model performance. One caveat is that any model likely represents a system of lower complexity than the real ocean. Also, when working with real observations, there are uncertainties associated with measurements, which is not accounted for in the model validation.

When trained with the synthetic dataset and used to predict the full NorESM model fields, ResNet-DIC achieves a global RMSE of $10.1 \pm 0.5 \mu\text{mol/kg}$ with a bias of $-0.45 \pm 0.12 \mu\text{mol/kg}$ across the 15-fold ensemble (figure 5). Performance varies strongly with depth. Errors are highest in the upper ocean above 150 m, where RMSE reaches $20.8 \pm 0.8 \mu\text{mol/kg}$, consistent with the greater variability of surface waters. In contrast, the deep ocean below 1500 m is reconstructed comparatively well, with RMSE values of $6.6 \pm 0.4 \mu\text{mol/kg}$ and low bias. Intermediate depths show moderate errors between these extremes. Across the 1990–2019 period, both RMSE and bias tend to increase toward the beginning and end of the time series, only partly coinciding with changes in training data density and potentially indicating temporal edge effects. Biases remain consistently negative and become more pronounced in later years.

Similar behaviour is observed across major ocean basins (figure 6), with basin-mean RMSE values generally between 6–9 $\mu\text{mol/kg}$ and modest negative biases. Further analysis from depth-separated median errors (figure 7) and mapped differences between model truth and predicted values (see figure A4) shows that the largest errors occur in coastal oceans and particularly in the Arctic Ocean. Elevated errors are also found in several coastal regions, including Central America, the western Indian Ocean, and semi-enclosed seas such as the Mediterranean. These regions are characterised by comparatively sparse training

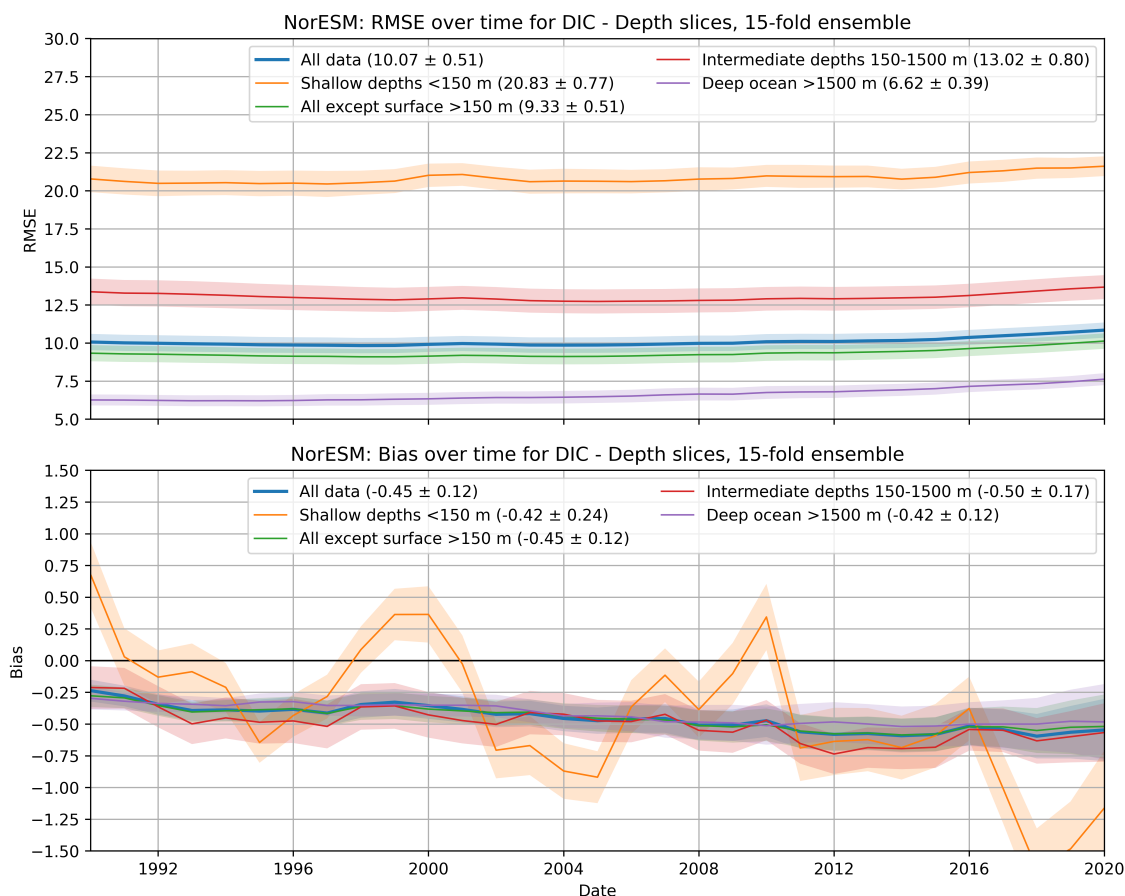


Figure 5. Errors (top panel) and bias (bottom panel) weighted by gridcell volume when predicting an annual mean of NorESM model output from 1990 to 2019. Plotted lines show different depth slices. Shading shows standard deviation over a 15-fold ensemble.

data coverage after quality-control filtering, especially in the Mediterranean Sea. The shallow ocean remains the main source of elevated and more variable errors, whereas deeper waters exhibit more stable prediction performance.

Results obtained with the ECCO-Darwin synthetic dataset show broadly similar overall performance. Global RMSE values are comparable to NorESM, with an overall RMSE of $9.6 \pm 0.9 \mu\text{mol/kg}$ and a continued tendency toward lower errors in the deep ocean than in shallower layers (figure A5). The average bias is $-0.50 \pm 0.14 \mu\text{mol/kg}$. However, ECCO-Darwin exhibits stronger temporal variability in both RMSE and bias, particularly across individual basins (figure A6). The Arctic Ocean shows elevated errors at both the beginning and end of the time series, while basin-scale biases vary in sign and magnitude through time, in contrast to the more consistently negative biases found for NorESM. Indian, Atlantic and Southern Ocean bias partly rises to positive values during the time series. These differences likely reflect both structural differences between the underlying ocean models and limitations inherent to the synthetic dataset construction, including the matching of discrete observations to gridded model output and also possible model drift in ECCO-Darwin, as a similar issue was encountered and the first three

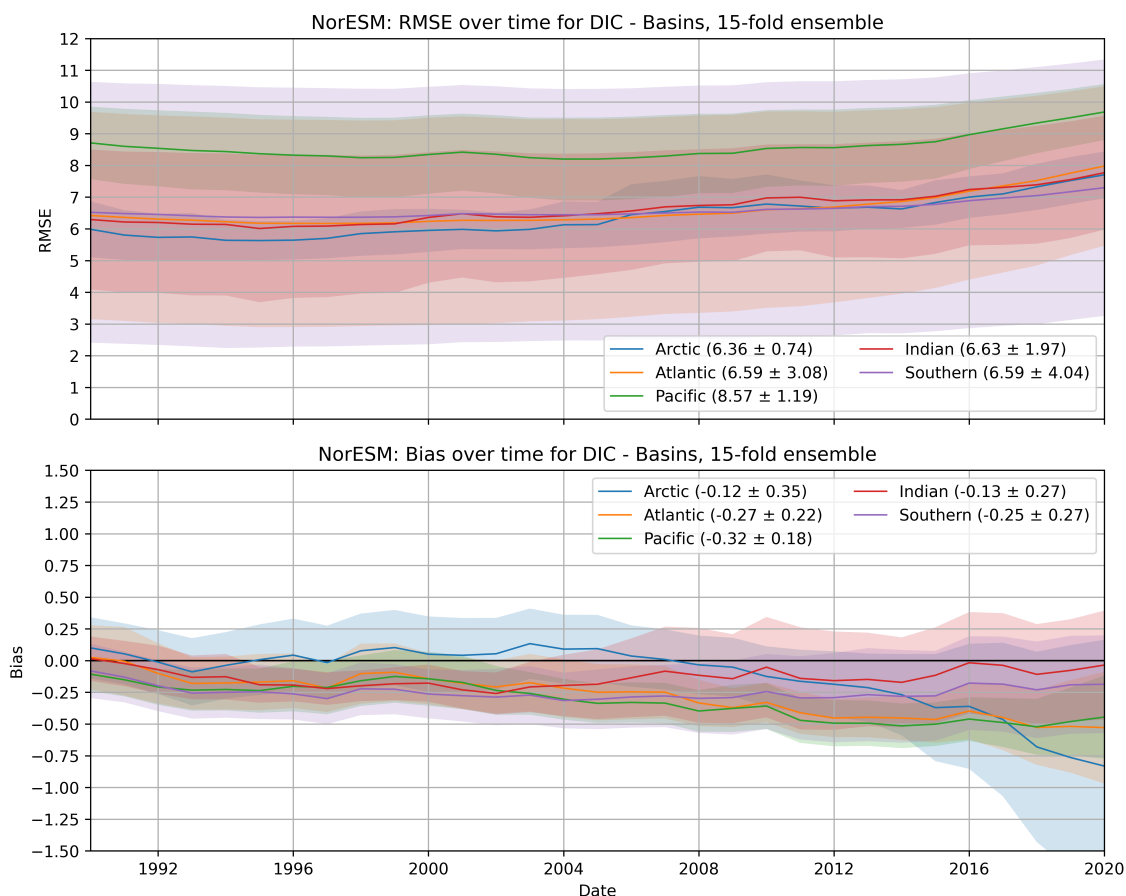


Figure 6. Errors (top panel) and bias (bottom panel) weighted by gridcell volume when predicting an annual mean of NorESM model output from 1990 to 2019. Plotted lines show different basins. Shading shows standard deviation over a 15-fold ensemble.

years discarded in the analysis of Carroll et al. (2022). Nevertheless, the two experiments yield a consistent overall picture: prediction skill is strongest in the deep ocean and weakest in shallow, highly variable, or sparsely sampled regions.

Factors resulting in higher errors for coastal and shelf seas include higher DIC variability, and smaller depth range resulting in fewer training data. In the Arctic Ocean, additional challenges arise from strong seasonal and spatial sampling biases, with observations concentrated in summer months and unevenly distributed between the Beaufort and Norwegian Seas. Nevertheless, we retain these regions in the analysis to avoid excluding areas of ecological and biogeochemical importance.

Despite often being identified as a particularly challenging region because of sparse observations (Gruber et al., 2019), the Southern Ocean is predicted comparatively well by ResNet-DIC in both synthetic validation experiments. Basin-scale RMSE values remain relatively low, especially in the ECCO-Darwin experiment with $3.6 \pm 0.1 \mu\text{mol/kg}$. One possible reason for this performance is that, despite sparse coverage, the Southern Ocean features large-scale, coherent circulation patterns like the Antarctic Circumpolar Current and fewer coastal or shelf regions, both of which reduce variability that typically challenges

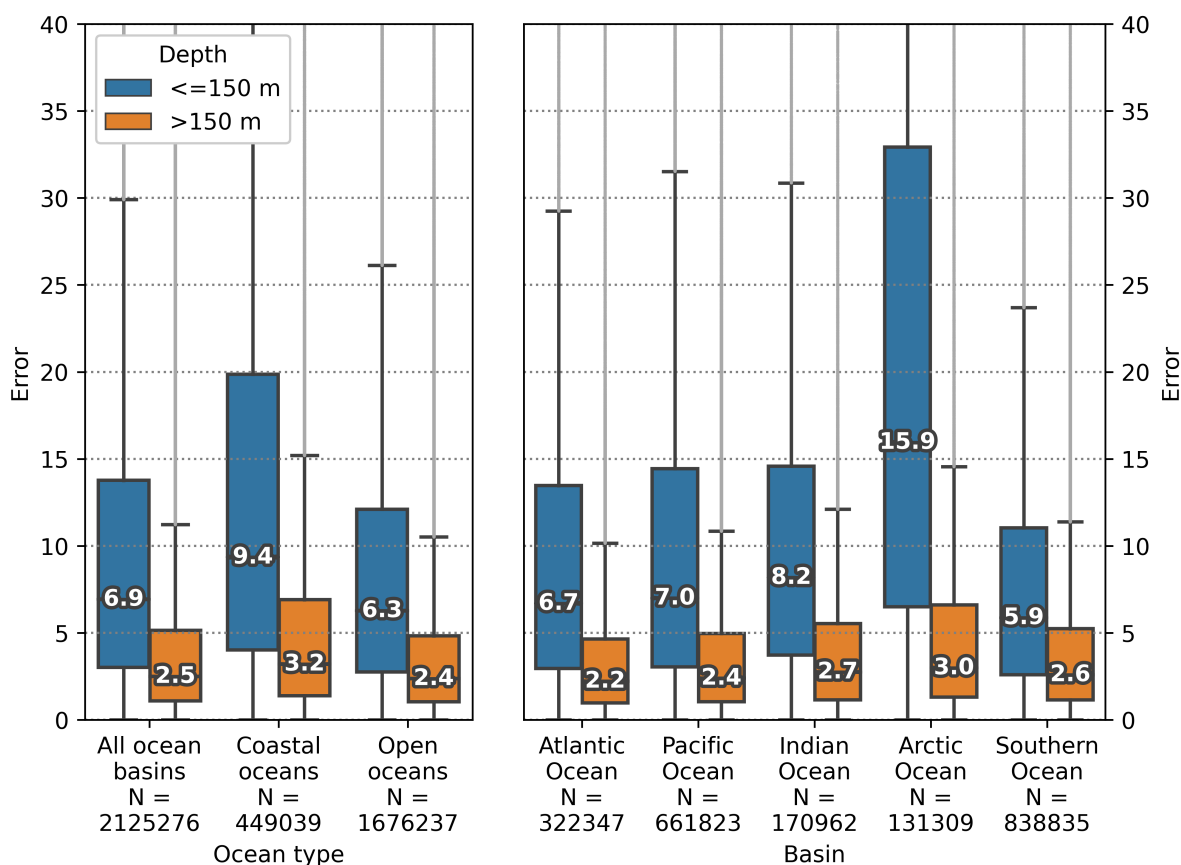


Figure 7. Error distribution for the year 2016 when predicting the ECCO-Darwin model comparing all, coastal, and open oceans (left panel); and different RECCAP2 basins (right panel). Shallow ocean layers above 150 m shown in blue, deep layers below 150 m shown in orange. Note the high errors in shallow coastal areas in general and specifically the shallow Arctic Ocean, but also the relatively low errors in the Southern Ocean. Labels show median errors.

prediction. Furthermore, DIC variability in this region is more tightly linked to physical drivers such as upwelling, which may be captured well enough by our predictor variables temperature and salinity.

300 We also evaluated the sensitivity of the validation results to the partitioning strategy used to construct the training and validation datasets. In addition to splitting by complete vertical profiles, we tested a split based on entire research cruises. Results from the synthetic-data experiments (Figure A7) indicate only minor differences between the two approaches. Profile-based splitting yields slightly lower errors in the major ocean basins, whereas cruise-based splitting performs somewhat better in semi-enclosed seas (masked in the RECCAP2 basin masks shown in Figure A7). These differences largely compensate,
 305 resulting in very similar global error metrics across both partitioning strategies. These findings also suggest that the detrimental effects of vertical-profile splitting may be less severe than inferred from the observational validation results alone.



3.2 Validation with independent time series

Other independent datasets well suited to validation are time series observations. Time series provide another challenge for the prediction, as they are sampled at much shorter time intervals and errors due to short-timescale variability tend to dominate our performance metrics. The data on which ResNet-DIC is trained do not include such sampling – long-running time series datasets are not included in GLODAP to avoid the introduction of biases towards a single location (Olsen et al., 2016) – so the model is not expected to closely reproduce seasonality. We use ResNet-DIC to predict the Hawaii Ocean Time-Series (HOT) (Dore et al., 2009), the Bermuda Atlantic Time-Series (BATS) (Bates et al., 2014), and the LDEO Drake Passage Surface Water CO₂ Time-Series Program (Takahashi et al., 2018) and compare the measured values with the predicted ones. The prediction of HOT DIC data achieves a RMSE of $8.54 \pm 1.1 \mu\text{mol/kg}$ with a bias of $0.8 \pm 1.3 \mu\text{mol/kg}$. Predicting BATS results in a RMSE of $9.5 \pm 0.8 \mu\text{mol/kg}$ with a bias of $0.94 \pm 0.7 \mu\text{mol/kg}$. Predicting the Drake Passage time series proves more challenging, with a RMSE of $23.1 \pm 1.0 \mu\text{mol/kg}$ with a bias of $-4.4 \pm 1.1 \mu\text{mol/kg}$. However, these results are overall encouraging. Compared to MOBO-DIC, for which the same time series were predicted in Keppler et al. (2023), the predictions of ResNet-DIC improved the RMSE by half and the bias by more than two thirds.

As suggested by the more challenging prediction of the Drake Passage time series, which contains only surface data, these reconstructions are less suited for the near-surface. The surface ocean is harder to predict due to its strong variability. Furthermore, less than 10% of all vertical profiles in GLODAPv2.2023 start at a depth of 0 m. Most have a minimum depth of 5 metres, likely to ensure full submergence of the sensors. Here, the addition of satellite data would likely improve results.

3.3 Uncertainty estimation

To estimate uncertainty in the ResNet-DIC reconstructions, we follow the framework of Keppler et al. (2023) and divide the uncertainty into three components.

1. Measurement uncertainty, the uncertainty in the underlying DIC observations. Since GLODAP does not explicitly provide measurement uncertainties for each observation, we adopt the basin-scale estimate of $4 \mu\text{mol/kg}$ reported for GLODAPv2.2023 (Lauvset et al., 2024). This value is more conservative than the $2.4 \mu\text{mol/kg}$ estimate used in Keppler et al. (2023) based on GLODAPv2.2021.
2. Representation uncertainty, a measure of how well the GLODAP sampling captures variability in the global ocean. Following Gregor and Gruber (2021) and Keppler et al. (2023), we adopt a value of $16 \mu\text{mol/kg}$, based on estimates for total alkalinity. While data coverage differs somewhat between DIC and total alkalinity, particularly prior to 2000, we retain this estimate as a conservative approximation.
3. Prediction uncertainty, which we estimate as the standard deviation of our 15-fold machine learning ensemble.

The resulting global mean combined uncertainty is $17.1 \mu\text{mol/kg}$ and is dominated by the representation uncertainty component. The spatial distribution of the uncertainty is shown in figure 8. The highest uncertainty values are in similar areas



as those affected by large errors in the validation with independent datasets and model data: shallow coastal areas in general, semi-enclosed seas, the Arctic shelves, and coastal Central America. One exception is the western Indian Ocean, which has low uncertainty but high errors in the validation (see figure A4). Higher uncertainties are also found near the contours of increased DIC in the central eastern Pacific Ocean (see section 4.1), suggesting that the position of these increased values has a high variability between ensemble members.

As in Keppler et al. (2023) the bulk of these values comes from the representation uncertainty. The variability in the uncertainty distribution comes solely from our prediction uncertainty, as all other components are constant. For the same depth range as MOBO-DIC, ResNet-DIC's uncertainty is $17.37 \mu\text{mol/kg}$, which remains slightly lower than MOBO-DIC's reported $18 \mu\text{mol/kg}$. This indicates that our prediction uncertainty is reduced compared to MOBO-DIC, even when adopting the updated, more conservative measurement uncertainty of $4 \mu\text{mol/kg}$.

4 Results and Discussion

4.1 DIC distribution

Similar to the spatial distribution of DIC in available DIC climatologies (Lauvset et al., 2016; Broullón et al., 2020), the ocean surface shows the lowest concentrations of DIC, especially shallow mid-latitudes (see figure 9). The Southern Ocean, along with the Atlantic and the Arctic Ocean has the highest surface concentrations. DIC concentrations generally increase with depth and range from $2000\text{--}2100 \mu\text{mol/kg}$ at the surface to $2200\text{--}2400 \mu\text{mol/kg}$ in greater depths. DIC concentrations are also high in upwelling areas, for example in the Eastern Equatorial Pacific, in areas affected by the Humboldt Current and the California Current, the eastern boundary upwelling system of West Africa and the equatorial upwelling region in the Atlantic. For these regions, the values are even higher than found in the NNGv2LDEO climatology (Broullón et al., 2020). The highest values can be found in the deep Central and North Pacific and Indian Oceans. Plots of zonal means (see figure 10) further reveal the depth extent of low DIC values in the mid latitudes, and the influence of the overturning circulation. The North Atlantic generally exhibits low DIC concentrations due to deep water formation and the sinking of recently ventilated waters. In contrast, higher DIC accumulation is observed in the interior of the North Pacific and North Indian Oceans, as these regions lie at the terminus of the lower limb of the overturning circulation (Talley, 2013). Furthermore, ventilation along sloping isopycnals of the Southern Ocean is evident in anomalous DIC extending towards the equator at depth in all three basins.

Hovmöller plots in figure 11 show the annual change of DIC over different basins. While there are short-term fluctuations in DIC accumulation in shallow and intermediate waters (up to 1500 m) over periods of 1-3 years, these variations are likely influenced by noise and should be interpreted with caution. Over longer timescales, however, a consistent increase in DIC accumulation rate becomes apparent. Deeper ocean layers below 1500 m also show this increased accumulation rate, with further acceleration especially in the Atlantic, Arctic and Southern Ocean. The deep Atlantic Ocean seems to undergo the most significant changes, from decreasing DIC accumulation until 2000 to a strong increase from 2016. However, this increase coincides with rising spread within the ensemble and therefore decreased confidence in the prediction. The figures also show

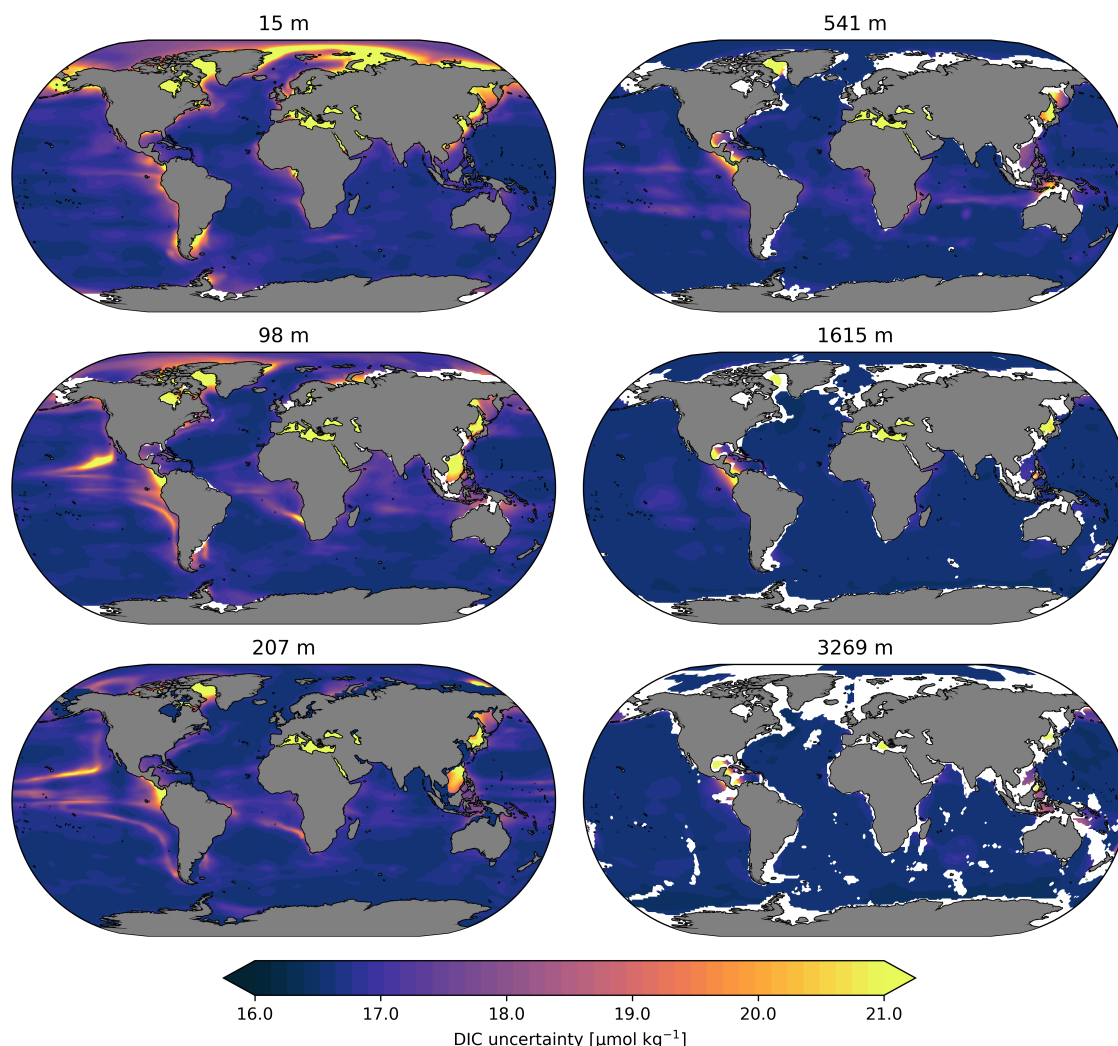


Figure 8. Spatial distribution of uncertainty for the DIC reconstruction. As other uncertainty components are constant, the variability shown here results from the prediction uncertainty: the standard deviation of the 15-fold machine learning ensemble.

370 bottom-intensified DIC accumulation in the deep ocean, which could arise from altered transport of e.g. North Atlantic Deep Water or Antarctic Bottom Water.

Overall, ResNet-DIC shows relatively good agreement with MOBO-DIC and climatologies if interpolated on the same spatiotemporal grid. Compared with the year-2002-normalised GLODAPv2.2016b climatology (Lauvset et al., 2016), ResNet-DIC shows elevated surface concentrations in the Southern and Arctic Oceans and larger discrepancies in the upper and intermediate layers up to 1000 m, whereas agreement improves in the deep ocean (figure A8). Relative to NNGv2LDEO, ResNet-DIC is generally consistent in deep waters but exhibits higher DIC concentrations in the upper ocean (figure A9). Compared with

375

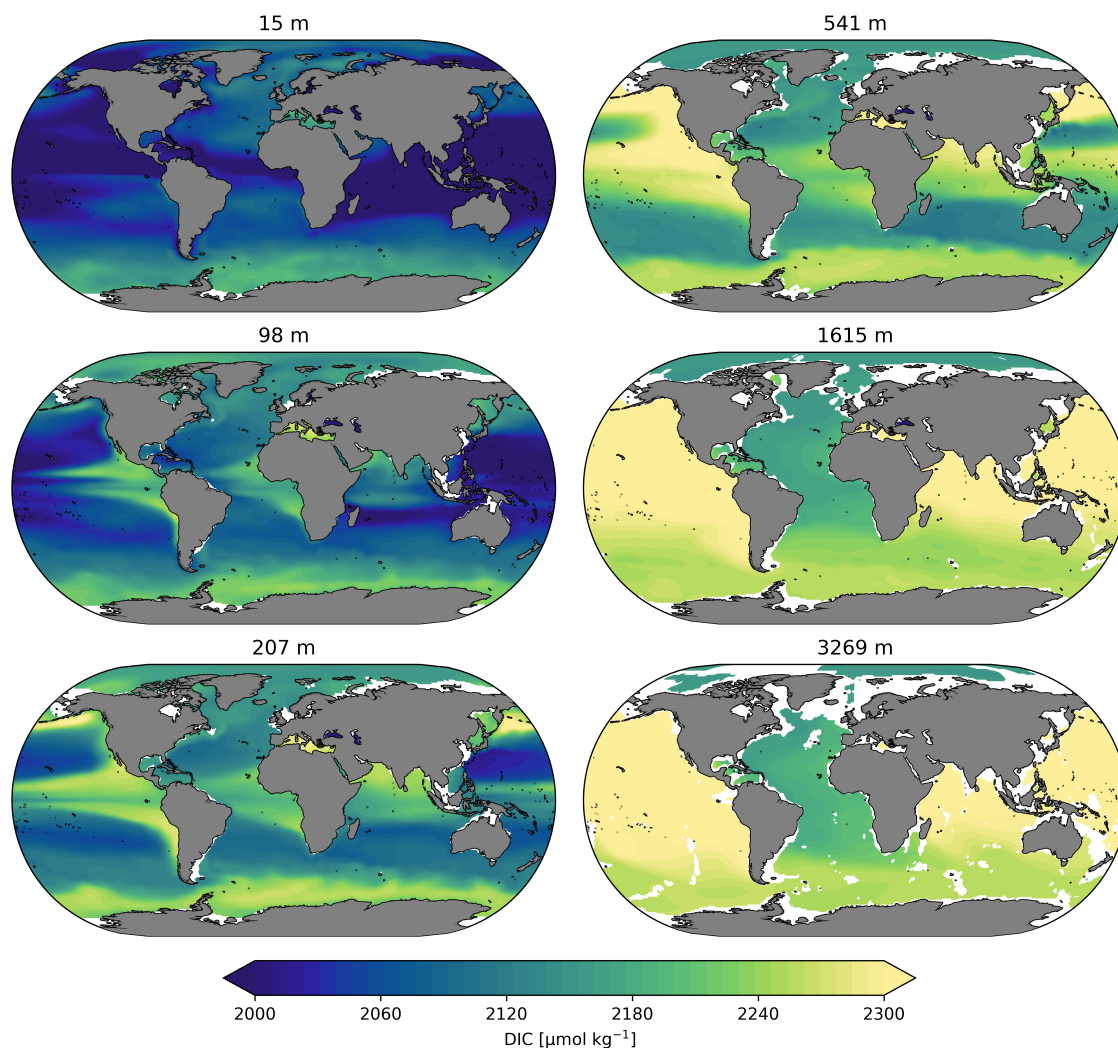


Figure 9. Temporal mean of the DIC reconstruction showing the global spatial distribution of DIC on a range of depth slices (number stated is the centre of the chosen EN4 depth bin).

MOBO-DIC, the largest differences occur in the Pacific and Indian Oceans, where ResNet-DIC generally predicts higher DIC concentrations, while the Atlantic Ocean is characterised by slightly lower values except in the upper 100–350 m (figure A10). The strongest regional discrepancy is found in the upper 150 m of the Arctic Ocean, where ResNet-DIC is more than 15 $\mu\text{mol kg}^{-1}$ lower than MOBO-DIC.

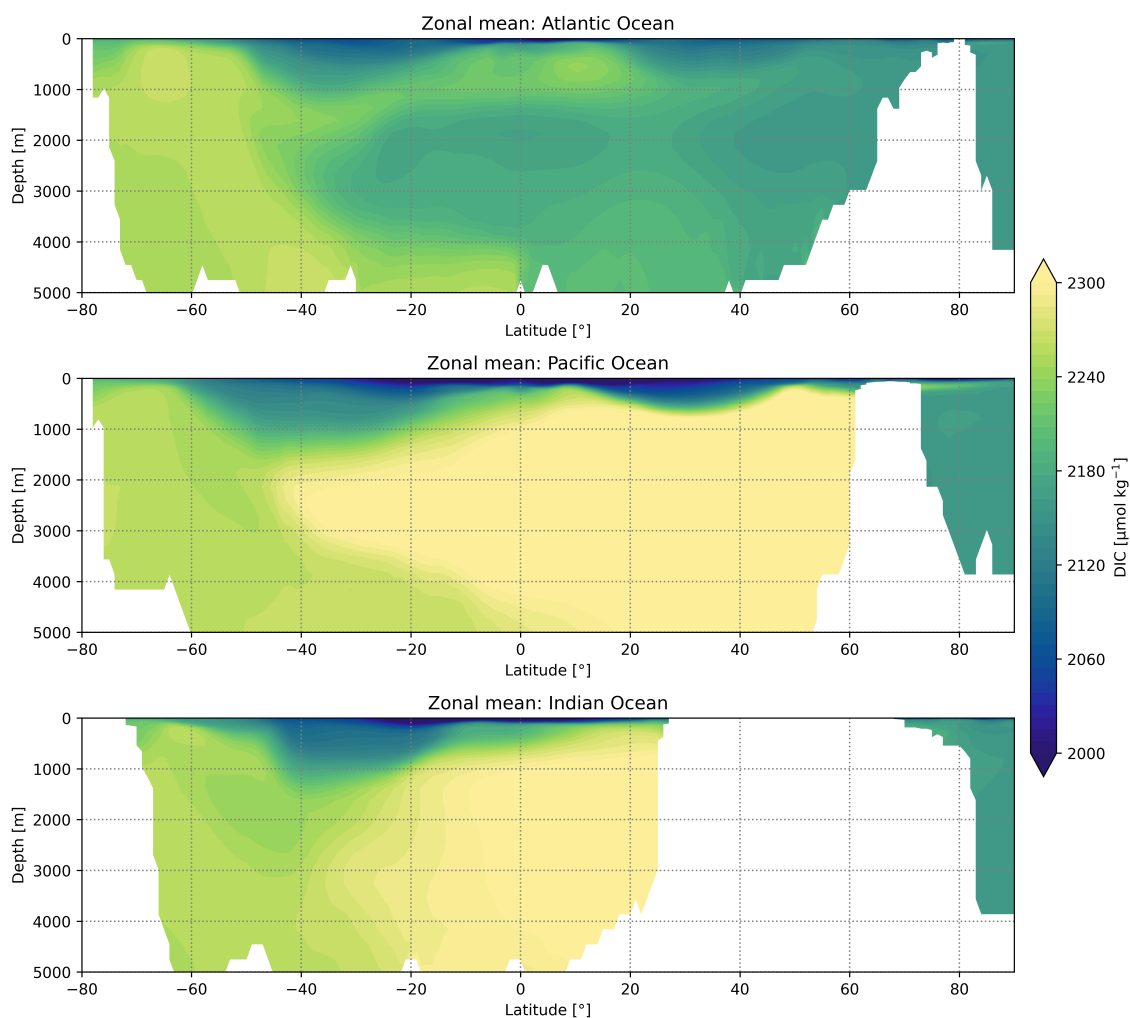


Figure 10. Temporal and zonal means of the DIC reconstruction showing the global spatial distribution of DIC across the major ocean basins.

4.2 Model-based carbon inventory

To assess whether a neural network trained on a sparse dataset can reconstruct carbon inventory changes, we apply the model validation framework described in Section 3.1 to predict the carbon inventory evolution from 1990 to 2019. Figures 12 and A11 show the reconstructed carbon inventory change for 15 ensemble members compared with the model truth of ECCO-Darwin and NorESM. Averaged over 1990 to 2019, the trend is predicted as 1.99 ± 0.12 Pg C/yr compared to the model truth of 2.16 Pg C/yr for NorESM and 1.38 ± 0.19 Pg C/yr compared to the model truth of 1.64 Pg C/yr for ECCO-Darwin (see middle panel in figures 12 and A11). The resulting differences of 0.25 Pg C/yr and 0.23 Pg C/yr, respectively, are consistent with the negative bias identified in Section 3.1. Carbon inventories, however, indicate that these differences likely reflect stronger

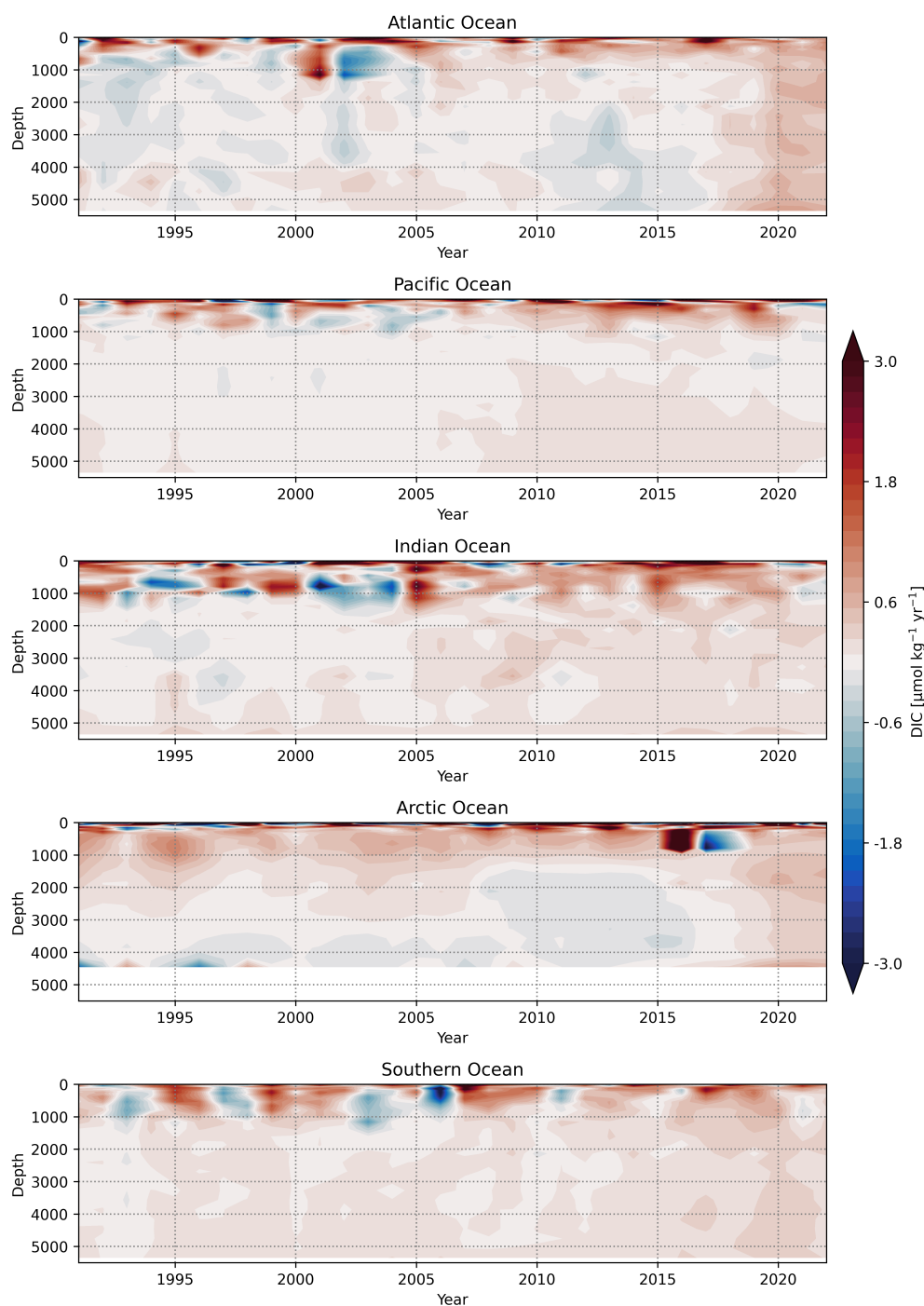


Figure 11. Hovmöller plots with depth against year showing the change in DIC per year.



underestimation in the early 1990s rather than a persistent offset throughout the full period. Afterwards, simulated trends are
 390 more closely tracked. When restricting the analysis to the MOBO-DIC period (2004–2019), for example, the predicted trend
 is 2.31 ± 0.21 Pg Pg C/yr compared with the model truth of 2.41 Pg Pg C/yr for NorESM, and 1.87 ± 0.24 Pg C Pg C/yr
 compared with the model truth of 1.92 Pg C/yr for ECCO-Darwin.

The spread of the ensemble members is strongest at the beginning and end of the time series. Except for the beginning of the
 time series, the standard deviation of the ensemble does not correlate well with the availability of training data (see lower panel
 395 in figures A11 and 12). Unexpectedly, around the year 2000 when training data availability is lowest, the standard deviation
 is particularly low. Assuming that ensemble spread is a measure of certainty in the predictions, this suggests that while the
 machine can compensate for years of low data by learning from other years, there may be additional uncertainties introduced
 through temporal edge effects (as discussed in section 3.1).

4.3 Observation-based carbon inventory

400 With interior ocean reconstructions that cover nearly the full depth of the ocean, we can calculate the global interior ocean
 carbon inventory and its rate of change. To calculate the thickness of each ocean layer represented by the EN4 grid we use the
 depth bounds included in the product. We find that EN4 does not represent accurate ocean depths, especially since the deepest
 depth grid cell has a maximum depth range of 5500 m. We correct for the thickness of the deepest wet grid cells using the
 GEBCO bathymetry dataset. Uncertainties for the carbon inventory are estimated by propagating the per-gridcell uncertainties
 405 calculated in section 3.3 in the same way as the inventory calculation.

Based on the model validation (section 3.1), semi-enclosed seas exhibit comparatively large biases due to limited high-
 quality training data coverage. To limit the influence of these poorly constrained regions on global carbon inventory estimates,
 predicted DIC values in semi-enclosed seas are replaced by the global mean depth profile during inventory calculation, while
 the gridded ResNet-DIC product remains unchanged. The impact of this correction is small and remains well within the
 410 estimated uncertainties, increasing the total inventory by about 2 Pg C and the trend by 0.02 Pg C, yr⁻¹.

To further assess the sensitivity of our results to uncertainties in the EN4 analysis dataset, we evaluated the predicted DIC
 inventory change using four different versions of EN4, each applying a distinct set of temperature and salinity corrections.
 The resulting carbon inventory trajectories (see figure A12) show only minor deviations across the four corrected datasets.
 While some differences in total carbon content are apparent - particularly before the year 2002 - the overall trends are highly
 415 consistent.

From 1990 to 2019, the total carbon content rises by 76 Pg C from a yearly and ensemble mean of $37,432 \pm 273$ Pg C to
 $37,508 \pm 274$ Pg C. Our estimates are higher than the $37,214 \pm 195$ Pg C estimated by Keppler et al. (2020), or the 37,484
 Pg C calculated from a 1992–2017 time mean of ECCO-Darwin (Carroll et al., 2022). Earlier published values, e.g. Falkowski
 et al. (2000), are of comparable magnitude but not intended to provide precision at this level.

420 Differences relative to published anthropogenic carbon uptake estimates are also within expected bounds. For 1994–2007, we
 estimate an increase of 25 Pg C compared to 29 ± 5 Pg C reported by Gruber et al. (2019). Note that these estimates represent
 anthropogenic carbon uptake and exclude contributions from natural carbon variability, so they are not directly equivalent to



total DIC storage change. The results are also lower for this time period compared to ECCO-Darwin, where the increase is 31 Pg C (Carroll et al., 2022). There is closer agreement in later decades (e.g. 1995–2018: 60 Pg C in this study versus 64 Pg C
 425 in ECCO-Darwin), indicating broadly consistent long-term behaviour, with differences primarily in decadal variability.

The change in the global DIC inventory with time for the machine learning ensemble is shown in Figure 13. Five-year averages are used to highlight the uptake trend, as even annualised uptake trends show a noisy signal (figure A1). Following Keppler et al. (2023), uncertainty in the trend is estimated from the ensemble spread. Overall, DIC accumulates in the interior ocean at an average rate of 2.64 ± 0.13 Pg C yr⁻¹ over 1990–2019, consistent with observational estimates of the anthropogenic
 430 carbon sink (2.7 ± 0.3 Pg C yr⁻¹ for 1990–2019; (Gruber et al., 2023)), noting that those estimates represent anthropogenic carbon uptake rather than total DIC storage change and exclude variations in the outgassing of C_{nat}.

The uptake rate is not temporally uniform. It increases from 1.61 ± 1.30 Pg C yr⁻¹ in 1990–1995 to 2.47 ± 0.82 Pg C yr⁻¹ in 1995–2000, followed by a pronounced slowdown to 0.89 ± 0.88 Pg C yr⁻¹ in 2000–2005 (see figure 13). This slowdown is reminiscent of the weakening of the ocean’s anthropogenic carbon sink that has been suggested by several surface
 435 pCO₂ reconstructions (e.g., Landschützer et al., 2016; Fay et al., 2021; Ford et al., 2024), although the onset of this change happens 2–3 years later in our reconstruction. Hovmöller plots (figure 11) indicate that this variability is primarily associated with intermediate-depth waters (300–1000 m), with coherent signals across the Atlantic, Indian, Southern, and parts of the Pacific Ocean. Following this period, uptake increases again to 2.66 ± 0.48 Pg C yr⁻¹ (2005–2010), 3.17 ± 0.57 Pg C yr⁻¹ (2010–2015), and 5.89 ± 0.89 Pg C yr⁻¹ (2015–2020). The late-period increase exceeds recent estimates of the anthropogenic
 440 carbon sink (3.1 ± 0.4 Pg C yr⁻¹ for the 2010s and 3.2 ± 0.4 Pg C yr⁻¹ for 2015–2024; (Friedlingstein et al., 2026)), but is associated with increased ensemble spread after 2016, indicating higher reconstruction uncertainty.

The accumulation rate accelerates from the reinvigoration after the slowdown in the early 2000s until 2020. This acceleration is likely due to the increase of CO₂ in the atmosphere (Keppler et al., 2023). However, there is a drastic decrease in the available training data from 2018 in GLODAPv2.2013, and fewer good EN4 temperature and salinity profiles in EN4 from 2019 (see
 445 figure A1), so the high rates of change near the end of the record must be viewed with caution. That the machine learning model is faced with increased uncertainty during those times is also represented in the larger ensemble spread at the end of the time series (see figure 13). However, the acceleration has also been observed in other studies, for example for a reconstruction of anthropogenic carbon in the Labrador Sea (Raimondi et al., 2021) or a reconstruction of DIC the Irminger Sea (Fröb et al., 2018). Hovmöller plots (see figure 11) show that the increase around 2015 is driven by the deep Pacific Ocean, Indian and
 450 Southern Ocean. The strongest increase after 2015 is mainly driven by the Atlantic, Southern, and Arctic Oceans.

The high variability of the uptake trend on annual time scales (see figure A1) indicates that the reconstruction is most likely unreliable as an indicator of net uptake of carbon on such short time scales, which is why we display trends as 5-year averages in figure 13. It is further affected by periods of little training data in the early 1990s and from 2019, no training data after 2021, and lack of high quality temperature and salinity profiles, of which there are fewer before 2004 and after 2019 (see
 455 figure A1). Periods of reduced training data density also coincide with greater ensemble spread in recent times, amplifying apparent short-term fluctuations that are unlikely to represent real changes in accumulation. Furthermore, during the periods

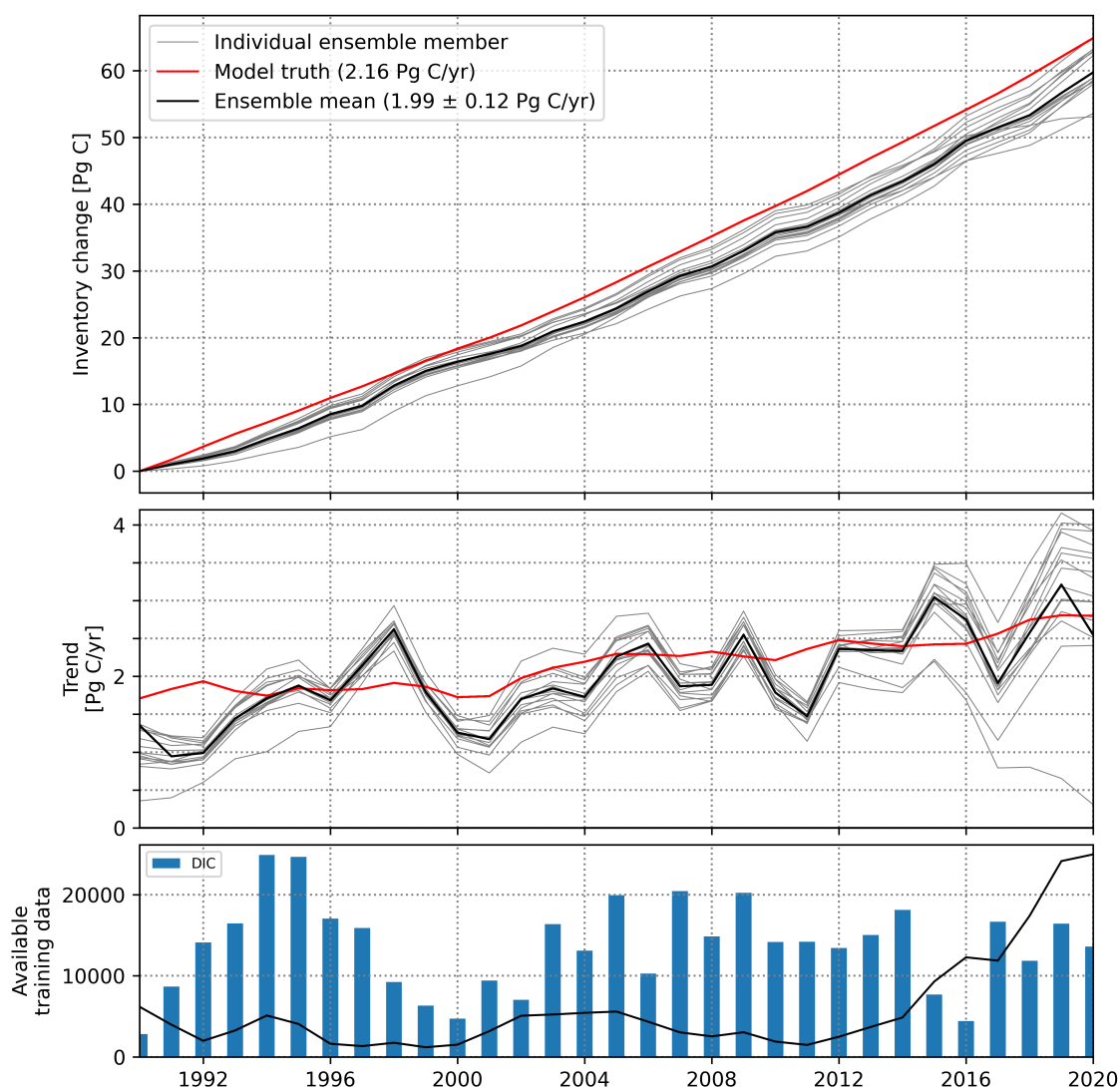


Figure 12. Annual mean of the carbon inventory change since 1990 (top panel) and uptake trend (middle panel, positive values indicating uptake of carbon) based on the model validation through training on a subset of NorESM, followed by a prediction of the full model fields. Model truth shown in red, individual ensemble members in grey, ensemble mean in black. Bottom panel shows the available training data compared to a line showing changing standard deviation within the ensemble.

of decreased accumulation rates, the magnitude of the trend, at times changing to net outgassing, is likely exaggerated. Such peaks in the trend are brief and likely not indicative of the real accumulation rate.

Figure A13 compares the ResNet-DIC reconstruction to the carbon inventory change of MOBO-DIC by interpolating it
 460 on the same spatiotemporal grid as MOBO-DIC. During the whole period 2004-2019 the carbon inventory rate of change

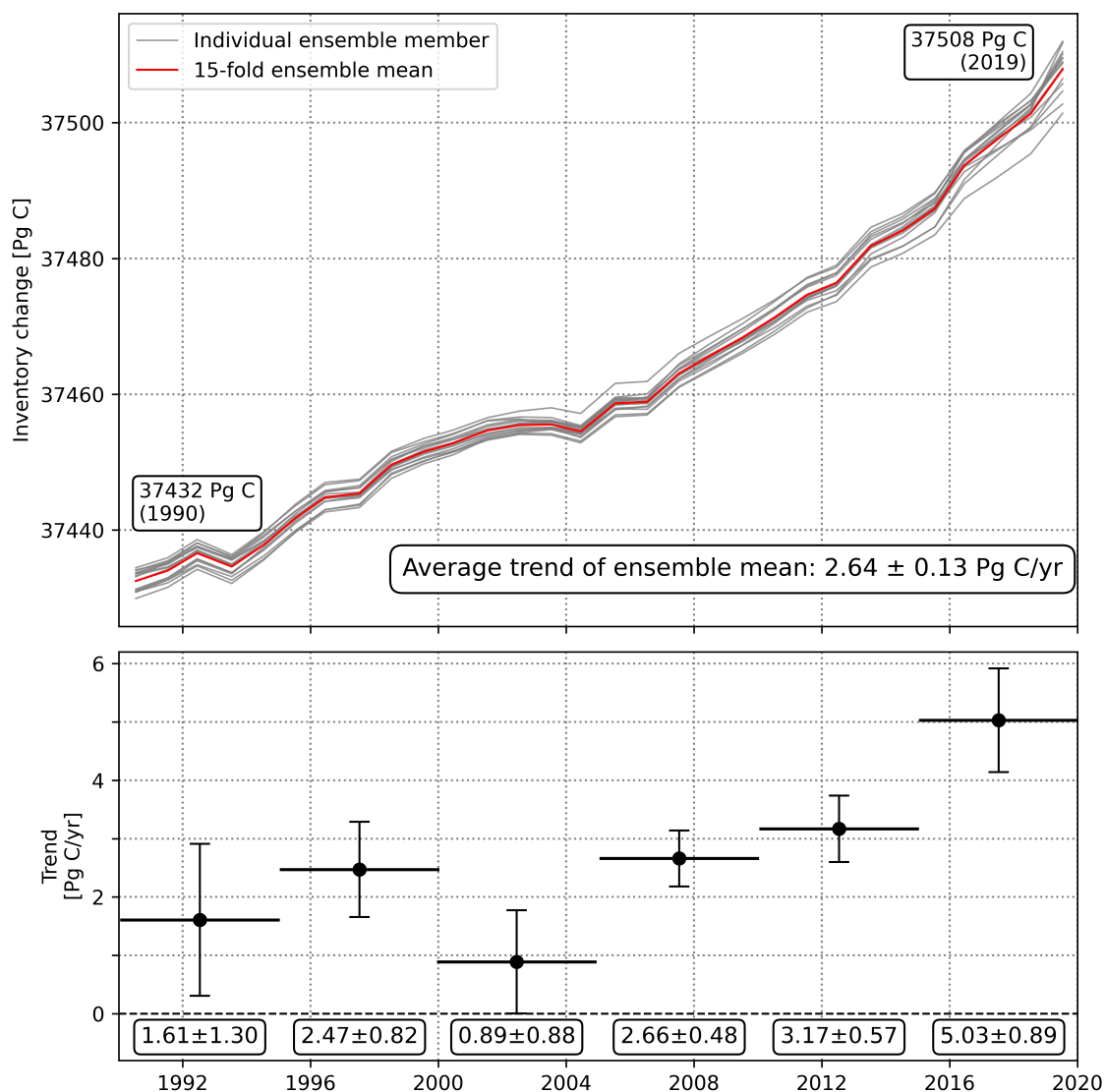


Figure 13. Annual mean of the total carbon inventory (top panel) and 5-year average of the uptake trend (bottom panel, positive values indicating uptake of carbon) for the 15-fold machine learning ensemble from 1990 to 2023. Ensemble mean shown in red, individual ensemble members in grey, standard deviation as the shaded area and 5-year averages as black horizontal bars with error standard deviation of the values inside each 5-year window as vertical error bars.

for the upper 1500 m from MOBO-DIC of 2.29 Pg C/yr is slower than from ResNet-DIC at 2.47 Pg C/year. The clearest divergence in accumulation rates occurs after 2014, when ResNet-DIC suggests an acceleration in carbon uptake, whereas MOBO-DIC indicates a slight weakening of the uptake rate. The temporal evolution in ResNet-DIC is broadly consistent with recent estimates reporting a strengthening ocean carbon sink in recent years (Friedlingstein et al., 2024). Interannual



variability is stronger than in MOBO-DIC, as expected given the use of smoothed climatologies in the MOBO-DIC training procedure. However, the increased variability does not overwhelm the large-scale temporal evolution, indicating that products trained on direct observations can retain a clear long-term signal despite higher noise levels. The distinct weakening of the DIC accumulation rate in the ResNet-DIC product in the early 2000s, in addition to the weakening of the anthropogenic carbon sink suggested by surface pCO₂ products, is not covered by MOBO-DIC's time range.

5 Summary and outlook

We reconstructed the spatiotemporal distribution of ocean interior DIC using the machine learning model ResNet-DIC. As the reconstruction is mapped on the EN4 analysis grid it has the same extent and resolution, allowing us to analyse global changes and changes across different basins.

Relying purely on discrete measurements during training has the potential to introduce noise, but also eliminates the introduction of biases and smoothness from interpolations if climatologies are used, as done in a previous reconstruction by Keppler et al. (2023). With ResNet-DIC at full depth we provide another observation-based estimate of the total DIC pool of $37,508 \pm 274$ Pg C in the year 2019. The ResNet-DIC product also shows a clear interannual and decadal variability in the DIC pool, which partly corroborates the variability suggested by surface pCO₂ products. A weakening and reinvigoration of the ocean's carbon sink in the early 2000s is reminiscent of the pattern of pCO₂ products used in the Global Carbon Budget (Friedlingstein et al., 2024) and a recent product by Ford et al. (2024), except it is delayed by 2-3 years. It is unclear if these observations are artifacts of the sparse observations on which this product is based. Encouraging results in validation with model and independent data, however, suggest that this product has some skill in reconstructing the spatiotemporal distribution. Unlike previous reconstructions, we included the Arctic Ocean and coastal seas. However, some of these areas showed high errors and high uncertainties, largely due to missing training data. The possible inclusion of regional datasets in future may further improve the reconstructions.

Beyond providing an estimate of the ocean interior DIC inventory, ResNet-DIC offers an independent observation-based reconstruction suitable for assessing long-term variability and trends in the ocean carbon sink. Current observational estimates reported in the Global Carbon Budget are derived from an ensemble of multiple products (Friedlingstein et al., 2026), reflecting the uncertainty associated with any individual reconstruction. We hope that products such as ResNet-DIC can contribute to the development of a broader suite of interior-ocean carbon reconstructions, helping to strengthen and independently validate the observational record of ocean carbon uptake.

Using this method with only the GLODAP dataset has the advantage of much faster neural network training. Training for a single ensemble member can be completed in as little as 1-2 hours and no high-performance computing resource is necessary. This means results can be more easily replicated and updated. Training data can be substituted with other datasets, new GLODAP releases, additional data not included in GLODAP, or different quality control measures can be applied prior to training. This is important as updated versions of carbon datasets are released, e.g. GLODAPv3, or adding separate datasets like SNAPO (Metzl et al., 2024).



We note that another reconstruction of Southern Ocean interior ocean DIC (Zemskova et al., 2022) used various biogeochemical parameters from remote sensing, predicted purely with data surface data and DIC from GLODAP data. This shows
 500 that the inclusion of satellite data adds valuable information for reconstructing DIC. However, we chose not to incorporate satellite data in this study to avoid significantly increasing the complexity and computational cost of the reconstruction process. Our current focus is on a fast, transparent reconstruction framework that remains consistent across depths and is easily extendable to new datasets. Nonetheless, integrating satellite data remains a promising avenue for future development.

There remain several caveats to address. The predictions are mapped onto the EN4 analysis product, which relaxes to
 505 climatology where observational coverage is sparse, and may therefore carry forward its smoothing, biases, and uncertainties. The uncertainty contained in EN4, specifically, when the machine learning model maps predictions on EN4, is not included in the uncertainty estimation. This is because they represent non-independent analysis errors without an accompanying correlation structure needed for consistent uncertainty combination (Good et al., 2013). Although training is based on discrete GLODAP measurements, the model maps to monthly averages on a regular grid with fixed depth levels, meaning reconstructions in
 510 data-poor regions or periods, particularly prior to the 1990s and after 2018, should be interpreted with caution. The choice of predictors also reflects a balance between information content and availability: while oxygen, silicate, and nitrate can improve reconstructions at depth, their sparse data coverage would likely increase global uncertainty, so this work relies primarily on temperature, salinity, and spatiotemporal variables.

While validation against numerical models provides a useful benchmark, it cannot capture all observational uncertainties
 515 because the synthetic dataset derived from monthly averages and fixed depth levels does not perfectly match the GLODAP dataset and does not include measurement errors. Furthermore, any model can be assumed to represent a different and likely simpler system than the real ocean.

Lastly, ResNet-DIC can be easily adapted to reconstruct other parameters for which the use of temperature and salinity records may be suitable, such as anthropogenic carbon, oxygen, or nutrients. As temperature and salinity constrain the ocean
 520 circulation which facilitates the transport of tracers, there is likely potential in applying a similar approach. We have conducted preliminary studies that demonstrate the feasibility of such adaptations, which will be elaborated upon in forthcoming publications.

Code and data availability. We provide the ensemble mean of ResNet-DIC as a netcdf file, the scaling factors for GEBCO bottom depths, and the code needed to replicate this analysis, on Zenodo at <https://doi.org/10.5281/ZENODO.14575968> (Ehmen et al., 2024).

525 *Data availability.* GLODAPv2.2023 (Lauvset et al., 2024) is publicly available from <https://glodap.info/index.php/merged-and-adjusted-data-product-v2-2023/>. EN4 analysis and profiles are publicly available from the UK Met Office Hadley Centre (<https://www.metoffice.gov.uk/hadobs/en4/>). ECCO-Darwin model data used for the validation are available for download from https://data.nas.nasa.gov/ecco/eccodata/lc_270/ecco_darwin_v5/output/monthly/.



Deaseasonalised monthly atmospheric CO₂ from Mauna Loa data can be obtained from NOAA at <https://gml.noaa.gov/ccgg/trends/>. REC-CAP2 basins used for analysing different basins is publicly available from <https://reccap2-ocean.github.io/regions/>. The data from the HOT, BATS, and Drake Passage time series are available at <http://hahana.soest.hawaii.edu/hot/hot-dogs/bextraction.html>, <http://bats.bios.edu/bats-data/>, and <https://www.nodc.noaa.gov/archive/arc0118/0171470/2.2/data/0-data/>.

The DIC climatology of by Broullón et al. (2020) is available at <https://doi.org/10.20350/digitalCSIC/10551>, the GLODAPv2.2016 climatology by Lauvset et al. (2016) at <https://glodap.info/index.php/mapped-data-product/> and the MOBO-DIC data product by Keppler et al. (2023) is available at <https://www.ncei.noaa.gov/data/oceans/ncei/ocads/metadata/0277099.html>. We use the GEBCO bathymetry dataset (GEBCO Bathymetric Compilation Group 2024, 2024) available at <https://www.gebco.net/>.

Appendix A: Introduction

In this appendix, we present additional information, analyses, and validation results that complement the main text, ordered broadly by their appearance in the manuscript. First, section describes additional analyses for predictor importance. Figure A1 shows the unaveraged uptake trend of DIC alongside the temporal availability of GLODAPv2.2023 observations and EN4 temperature and salinity profiles underlying the EN4 analysis product.

Additional validation statistics are provided for both observational and model-based validation experiments. Figures A2 and A7 show the distributions of RMSE and bias across different validation subsets, basins, and train/test split strategies for the GLODAP and model validation experiments, respectively. Figure A3 compares the data availability of the synthetic ECCO-Darwin training dataset with GLODAPv2.2023. Figure A4 illustrates spatial prediction errors from the ECCO-Darwin model validation, while figures A5 and A6 show the temporal evolution of volume-weighted RMSE and bias across depth layers and ocean basins.

Our reconstruction is further compared with existing climatologies and reconstructions. Figure A8 shows differences relative to the GLODAPv2.2016b climatology, figure A9 compares our reconstruction with the NNGv2LDEO climatology, and figure A10 presents differences relative to MOBO-DIC.

Figure A11 compares carbon inventory changes and uptake trends derived from the ECCO-Darwin model validation framework. Figure A12 compares carbon inventory changes obtained using different EN4 temperature and salinity correction schemes, and figure A13 provides a comparison of carbon inventory changes and uptake trends between ResNet-DIC and MOBO-DIC.

Lastly, we present predictor importance analyses using Shapley values, first comparing different predictor sets (figure A14) and subsequently examining the relative importance of predictors as a function of depth (figure A15).

A1 Predictor importance

One metric for analysing predictions from machine learning models is the model-agnostic Shapley value Shapley (1951); Slack et al. (2020) that we construct with the Python module SHAP Lundberg and Lee (2017). Through taking random subsets of multiple variables from existing data, Shapley values ensure that combinations of values stay realistic and take into account how variables influence each other. Mean absolute Shapley values for a given predictor variable show by how much on average



that variable changed the prediction. Shapley values for predictor variables used for the DIC reconstruction are shown in figure A14a. Temperature is by far the most influential predictor variable, followed by salinity, depth, location and atmospheric CO₂. The high influence of temperature is important for mapping the reconstruction on a global grid, as temperature has the most quality controlled measurements available, which decreases the introduced uncertainty further. If Shapley values are analysed over depth (see figure A15), the increasing influence of depth and location in intermediate to deep waters suggests that the composition of water masses given by temperature and salinity become less important with increasing depth. This is likely due to their low variability in deeper waters. Instead, the neural network is relying more on “memorised” spatial distribution.

To further assess the impact of excluding oxygen and dissolved inorganic nutrients, we trained separate neural networks with all predictors used for MOBO-DIC, substituting silicic acid with silicate, and adding further predictors location, depth and atmospheric CO₂. We then create Shapley values to analyse this combination of predictors (see figure A14b). Oxygen, silicate, and nitrate have a higher impact on the reconstruction and the absolute mean impact of both temperature and salinity is reduced from 58.4 to 36.2 $\mu\text{mol/kg}$ and 17.9 to 10.6 $\mu\text{mol/kg}$, respectively. This is partly because the addition of more predictor variables spreads the impact over several variables. Furthermore, the impact of oxygen and silicate increases with depth, while the impact of temperature, salinity and nitrate decreases (see figure S2). As discussed in section 2.2, this suggests that oxygen, silicate and nitrate strongly affect a DIC reconstruction, with oxygen and silicate especially impactful in the deeper ocean. However, this has to be weighed against the inferior data availability when mapping onto a global grid. The importance of adding location to the predictor variables found in section 2.2 is consistent with our findings using Shapley values: when excluding oxygen, phosphate, nitrate and silicate, the impact of location on DIC is 10.9 $\mu\text{mol/kg}$ or a 10% relative contribution on average (see figure A14a). If we add those variables back to the training, the importance of the location is drastically reduced to an average impact on the DIC prediction of 1.4 $\mu\text{mol/kg}$ or just over 1% relative contribution (see figure A14b). This is consistent with the view that location is acting similarly to these variables, enabling discrimination between water masses that are widely separated in location but have similar temperature and salinity properties.

When choosing predictor variables, there are more things to consider. Some predictor variables like atmospheric CO₂ can have low average Shapley values but have a significant impact on the reconstruction. Predicting without atmospheric CO₂ introduces a strong seasonal cycle in the reconstruction with unrealistically high changes in the global carbon inventory. This indicates that some variable to represent the time dimension, such as atmospheric CO₂ or decimal year, is needed to stabilise the seasonal variability of the prediction. Without a measure of time the neural network can mistake variability due to seasonality on shorter time scales with longer term changes in the climate system if they have the same magnitude.

Author contributions. T.E. designed and developed the machine-learning reconstruction, conducted the analyses, produced the figures, and wrote the original manuscript draft. N.M. and A.W. secured funding and contributed substantially to the study design, scientific framing, and manuscript development. N.M. provided continuous scientific feedback and performed additional analyses and validation of figures and results. All authors reviewed, edited, and approved the final manuscript.

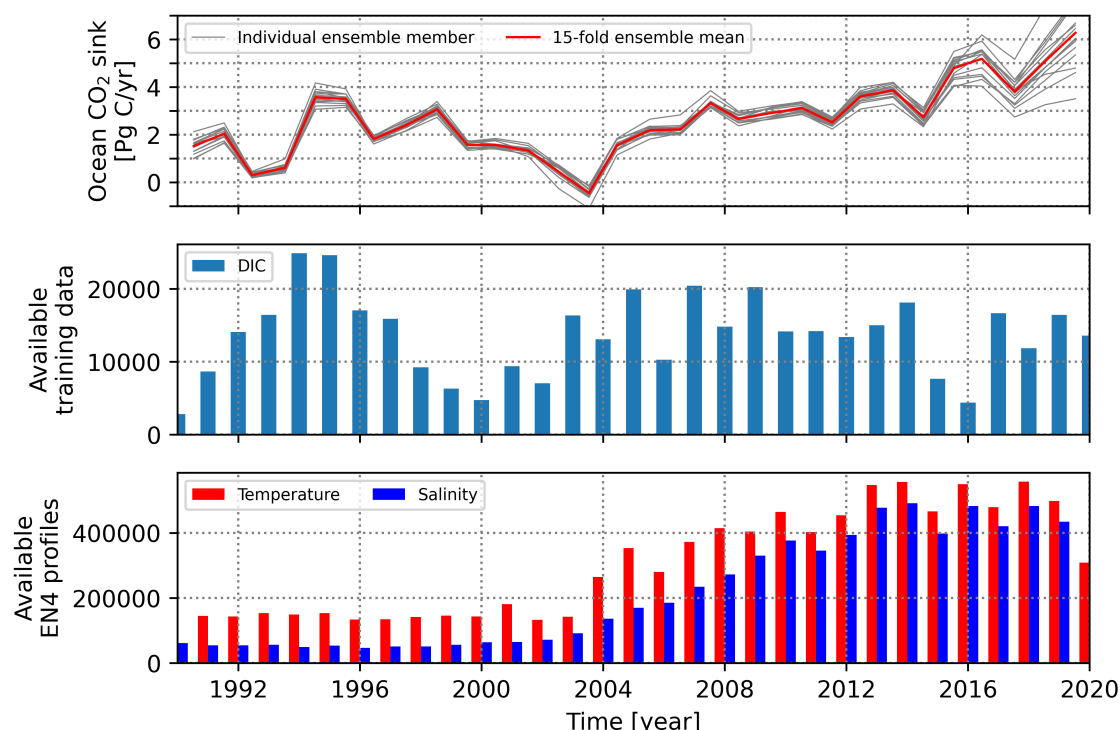


Figure A1. Top panel: annual uptake rates from the carbon inventory changes with individual ensemble members shown as black lines and the ensemble mean shown as a red line. Middle panel: data availability of GLODAPv2.2023 data over the predicted times. Bottom panel: number of good EN4 temperature (red) and salinity (blue) profiles on which the EN4 analysis is based (Good et al., 2013).

Competing interests. The authors declare no competing interests.

Acknowledgements. The authors acknowledge the GLODAPv2.2023 developers (Lauvset et al., 2024) for the training data, the EN4 developers (Good et al., 2013) for the use in our reconstruction and the ECCO-Darwin developers (Carroll et al., 2020) for the model outputs used in our model validation.

T.E., N.M. and A.W. have been supported by the Natural Environment Research Council (grant no. NE/W001543/1). T.E. and A.W. have also been supported by the European Union under grant agreement no. 101083922 (OceanICU) and UK Research and Innovation (UKRI) under the UK government's Horizon Europe funding guarantee (grant numbers 10054454, 10063673, 10064020, 10059241, 10079684, 10059012, 10048179). T.E. was additionally supported by the European Union's Horizon Europe programme under grant number 101130676 (NUBICOS) and 101094690 (EuroGO-SHIP). The authors declare that they have no conflicts of interest.

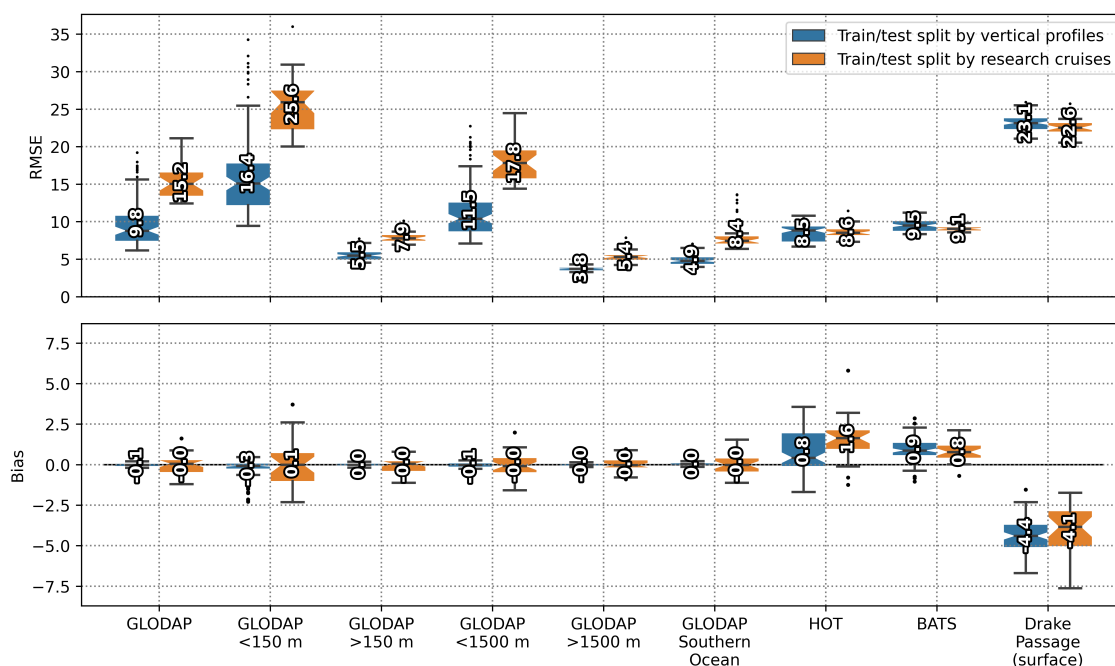


Figure A2. Distribution of RMSE (top panel) and bias (bottom panel) over different datasets and subsets. GLODAP refers to the validation dataset not seen during training, subsetted at different depths and the Southern Ocean. HOT, BATS and Drake Passage are time series datasets. The blue and orange bars refer to the different error metrics when performing the train/test split that creates the validation dataset: split either by complete vertical profiles (blue) or complete research cruises (orange). Statistics are the results of more than 30 ensemble members in each category. Labels show mean, boxes span the interquartile range, end caps extend to the smallest and largest values within 1.5 times the interquartile range and dots represent outliers.

References

- Bakker, D. C. E., Pfeil, B., Landa, C. S., Metzl, N., O'Brien, K. M., Olsen, A., Smith, K., Cosca, C., Harasawa, S., Jones, S. D., Nakaoka, S.-i., Nojiri, Y., Schuster, U., Steinhoff, T., Sweeney, C., Takahashi, T., Tilbrook, B., Wada, C., Wanninkhof, R., Alin, S. R., Balestrini, C. F., Barbero, L., Bates, N. R., Bianchi, A. A., Bonou, F., Boutin, J., Bozec, Y., Burger, E. F., Cai, W.-J., Castle, R. D., Chen, L., Chierici, M., Currie, K., Evans, W., Featherstone, C., Feely, R. A., Fransson, A., Goyet, C., Greenwood, N., Gregor, L., Hankin, S., Hardman-Mountford, N. J., Harlay, J., Hauck, J., Hoppema, M., Humphreys, M. P., Hunt, C. W., Huss, B., Ibáñez, J. S. P., Johannessen, T., Keeling, R., Kitidis, V., Körtzinger, A., Kozyr, A., Krasakopoulou, E., Kuwata, A., Landschützer, P., Lauvset, S. K., Lefèvre, N., Lo Monaco, C., Manke, A., Mathis, J. T., Merlivat, L., Millero, F. J., Monteiro, P. M. S., Munro, D. R., Murata, A., Newberger, T., Omar, A. M., Ono, T., Paterson, K., Pearce, D., Pierrot, D., Robbins, L. L., Saito, S., Salisbury, J., Schlitzer, R., Schneider, B., Schweitzer, R., Sieger, R., Skjelvan, I., Sullivan, K. F., Sutherland, S. C., Sutton, A. J., Tadokoro, K., Telszewski, M., Tuma, M., van Heuven, S. M. A. C., Vandemark, D., Ward, B., Watson, A. J., and Xu, S.: A multi-decade record of high-quality fCO₂ data in version 3 of the Surface Ocean CO₂ Atlas (SOCAT), *Earth System Science Data*, 8, 383–413, <https://doi.org/10.5194/essd-8-383-2016>, 2016.

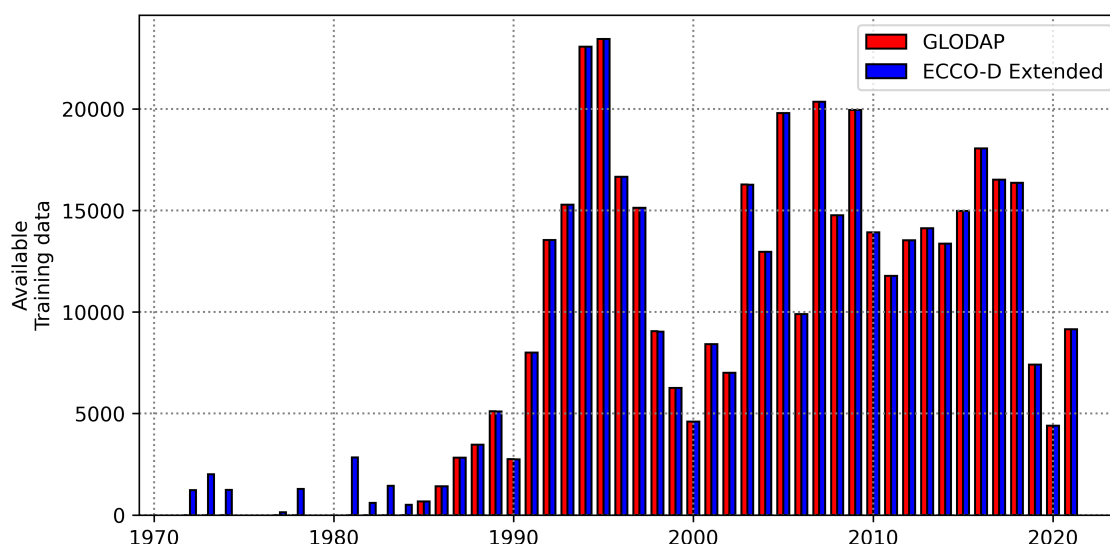


Figure A3. Comparison of number of datapoints in subsetted ECCO-Darwin synthetic dataset to datapoints in GLODAPv2.2023 when only "good" flags of both WOCE and GLODAP QC are retained. The range of ECCO-Darwin from 1985-2021 ensures that most GLODAP datapoints have a temporal match in ECCO-Darwin. As of GLODAPv2.2023 there are no datapoints more recent than 2021 and prior to 1985 there are few datapoint with good quality control flags.

- Bates, N., Astor, Y., Church, M., Currie, K., Dore, J., Gonaález-Dávila, M., Lorenzoni, L., Muller-Karger, F., Olafsson, J., and Santa-Casiano, M.: A Time-Series View of Changing Ocean Chemistry Due to Ocean Uptake of Anthropogenic CO₂ and Ocean Acidification, *Oceanography*, 27, 126–141, <https://doi.org/10.5670/oceanog.2014.16>, 2014.
- Boyer, T. P., García, H. E., Locarnini, R. A., Zweng, M. M., Mishonov, A. V., Reagan, J. R., Weathers, K. A., Baranova, O. K., Paver, C. R., Seidov, D., and Smolyar, I. V.: World Ocean Atlas 2018, NOAA National Centers for Environmental Information. Dataset., <https://www.ncei.noaa.gov/archive/accession/NCEI-WOA18>, 2018.
- Broullón, D., Pérez, F. F., Velo, A., Hoppema, M., Olsen, A., Takahashi, T., Key, R. M., Tanhua, T., Santana-Casiano, J. M., and Kozyr, A.: A global monthly climatology of oceanic total dissolved inorganic carbon: a neural network approach, *Earth System Science Data*, 12, 1725–1743, <https://doi.org/10.5194/essd-12-1725-2020>, 2020.
- Carroll, D., Menemenlis, D., Adkins, J. F., Bowman, K. W., Brix, H., Dutkiewicz, S., Fenty, I., Gierach, M. M., Hill, C., Jahn, O., Landschützer, P., Lauderdale, J. M., Liu, J., Manizza, M., Naviaux, J. D., Rödenbeck, C., Schimel, D. S., der Stocken, T. V., and Zhang, H.: The ECCO-Darwin Data-Assimilative Global Ocean Biogeochemistry Model: Estimates of Seasonal to Multidecadal Surface Ocean CO₂ and Air-Sea CO₂ Flux, *Journal of Advances in Modeling Earth Systems*, 12, <https://doi.org/10.1029/2019ms001888>, 2020.
- Carroll, D., Menemenlis, D., Dutkiewicz, S., Lauderdale, J. M., Adkins, J. F., Bowman, K. W., Brix, H., Fenty, I., Gierach, M. M., Hill, C., Jahn, O., Landschützer, P., Manizza, M., Mazloff, M. R., Miller, C. E., Schimel, D. S., Verdy, A., Whitt, D. B., and Zhang, H.: Attribution of Space-Time Variability in Global-Ocean Dissolved Inorganic Carbon, *Global Biogeochemical Cycles*, 36, <https://doi.org/10.1029/2021gb0007162>, 2022.

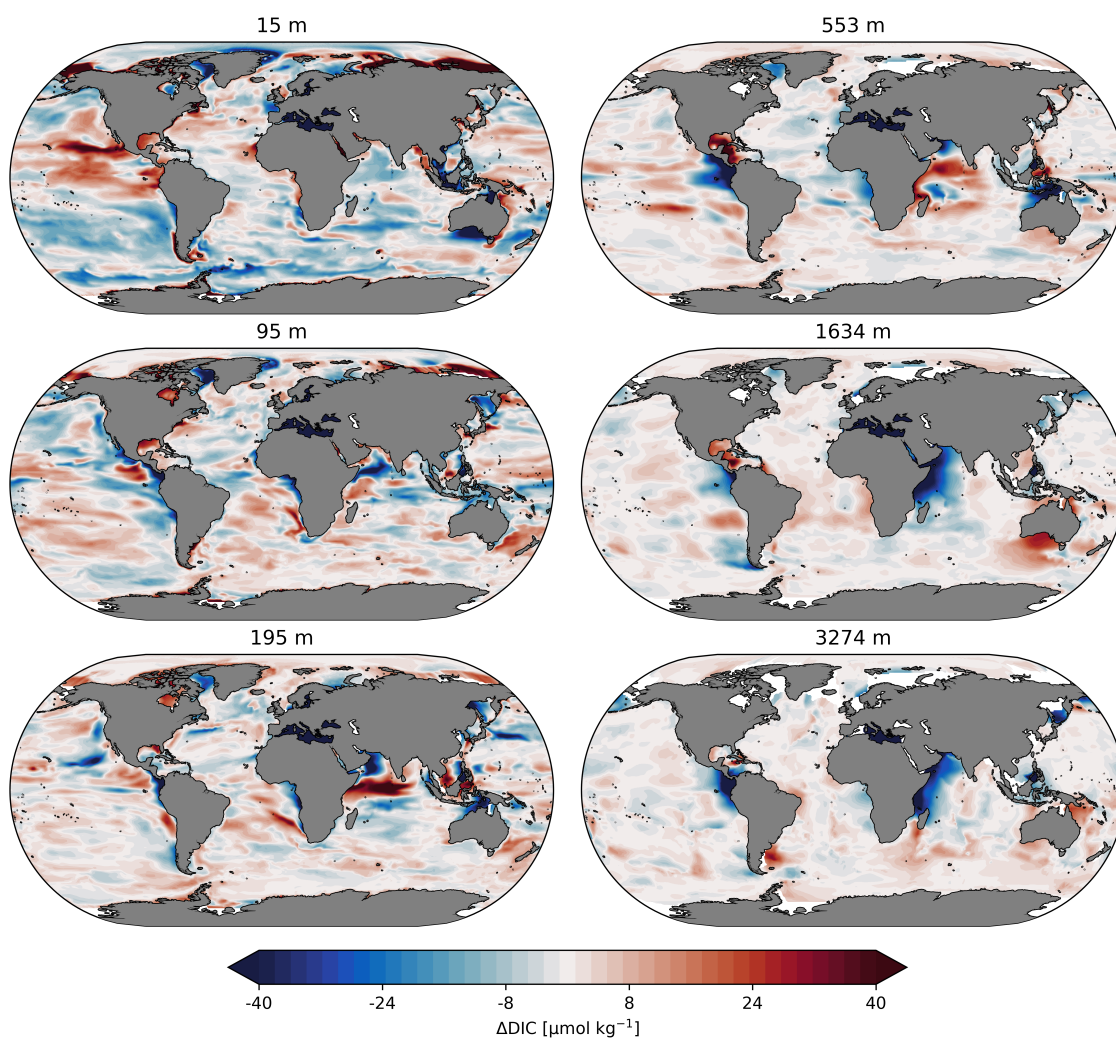


Figure A4. Difference between prediction and true values of ECCO-Darwin model data for the year 2016. Largest errors are mostly found in coastal areas and semi-enclosed seas.

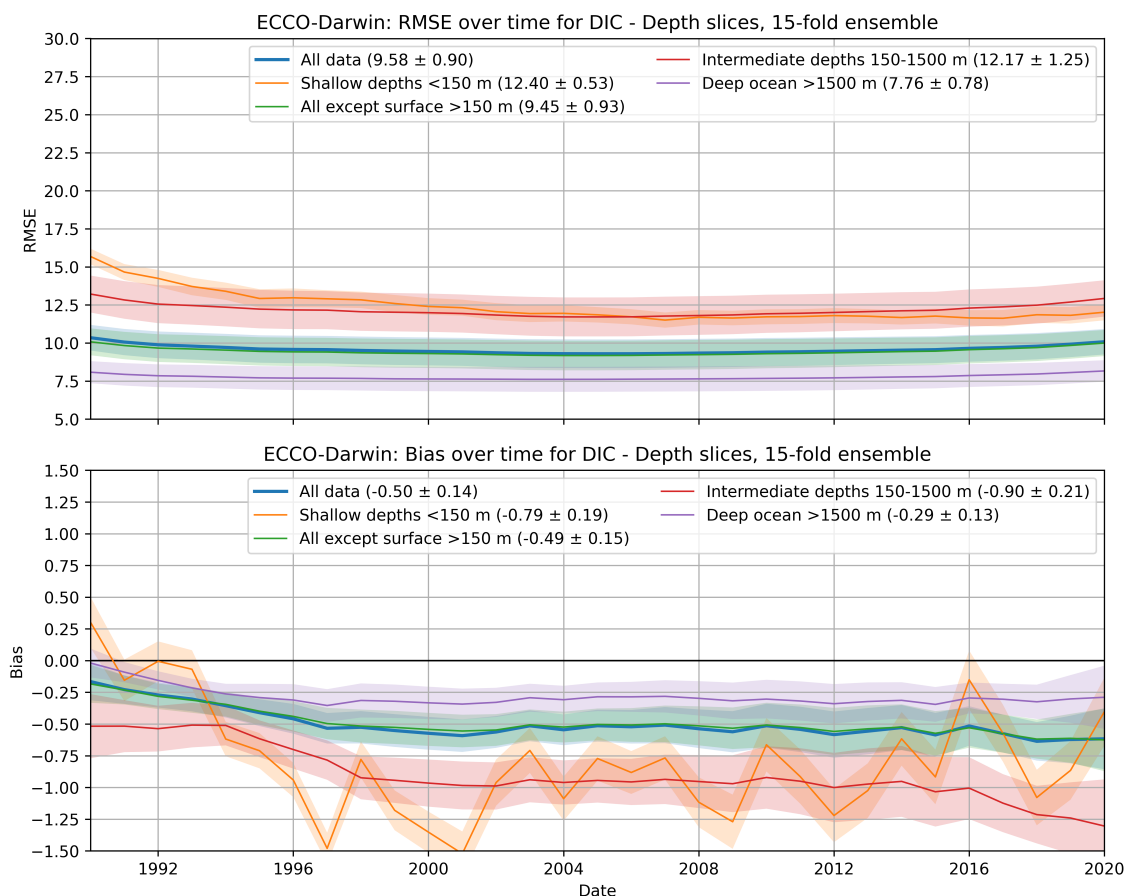


Figure A5. Errors (top panel) and bias (bottom panel) weighted by gridcell volume when predicting an annual mean of ECCO-Darwin model output from 1990 to 2019. Plotted lines show different depth slices. Shading shows standard deviation over a 15-fold ensemble.

- Carter, B. R., Schwinger, J., Sonnerup, R., Fassbender, A. J., Sharp, J. D., Dias, L. M., and Sandborn, D. E.: Tracer-based Rapid Anthropogenic Carbon Estimation (TRACE), *Earth System Science Data*, 17, 3073–3088, <https://doi.org/10.5194/essd-17-3073-2025>, 2025.
- DeVries, T. and Weber, T.: The export and fate of organic matter in the ocean: New constraints from combining satellite and oceanographic tracer observations, *Global Biogeochemical Cycles*, 31, 535–555, <https://doi.org/10.1002/2016gb005551>, 2017.
- 635 DeVries, T., Le Quéré, C., Andrews, O., Berthet, S., Hauck, J., Ilyina, T., Landschützer, P., Lenton, A., Lima, I. D., Nowicki, M., Schwinger, J., and Séférian, R.: Decadal trends in the ocean carbon sink, *Proceedings of the National Academy of Sciences*, 116, 11 646–11 651, <https://doi.org/10.1073/pnas.1900371116>, 2019.
- Doney, S. C., Fabry, V. J., Feely, R. A., and Kleypas, J. A.: Ocean Acidification: The Other CO₂ Problem, *Annual Review of Marine Science*, 1, 169–192, <https://doi.org/10.1146/annurev.marine.010908.163834>, 2009.
- 640 Dore, J. E., Lukas, R., Sadler, D. W., Church, M. J., and Karl, D. M.: Physical and biogeochemical modulation of ocean acidification in the central North Pacific, *Proceedings of the National Academy of Sciences*, 106, 12 235–12 240, <https://doi.org/10.1073/pnas.0906044106>, 2009.

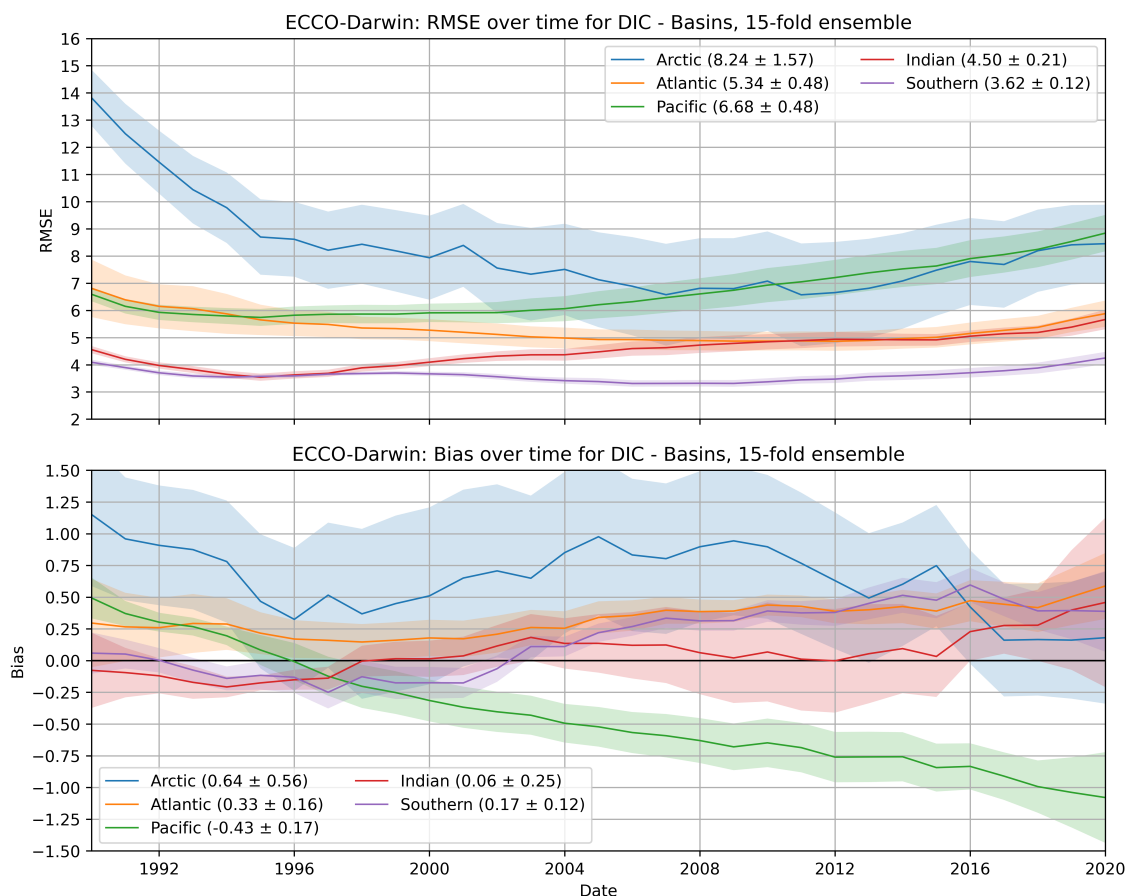


Figure A6. Errors (top panel) and bias (bottom panel) weighted by gridcell volume when predicting an annual mean of ECCO-Darwin model output from 1990 to 2019. Plotted lines show different basins. Shading shows standard deviation over a 15-fold ensemble.

Ehmen, T., Mackay, N., and Watson, A.: Data product and code for: Spatiotemporal Distribution of Dissolved Inorganic Carbon in the Global Ocean Interior - Reconstructed through Machine Learning, <https://doi.org/10.5281/ZENODO.14575968>, 2024.

645 Fabry, V. J., Seibel, B. A., Feely, R. A., and Orr, J. C.: Impacts of ocean acidification on marine fauna and ecosystem processes, *ICES Journal of Marine Science*, 65, 414–432, <https://doi.org/10.1093/icesjms/fsn048>, 2008.

Falkowski, P., Scholes, R. J., Boyle, E., Canadell, J., Canfield, D., Elser, J., Gruber, N., Hibbard, K., Höglberg, P., Linder, S., Mackenzie, F. T., Moore III, B., Pedersen, T., Rosenthal, Y., Seitzinger, S., Smetacek, V., and Steffen, W.: The Global Carbon Cycle: A Test of Our Knowledge of Earth as a System, *Science*, 290, 291–296, <https://doi.org/10.1126/science.290.5490.291>, 2000.

650 Fay, A. R., Gregor, L., Landschützer, P., McKinley, G. A., Gruber, N., Gehlen, M., Iida, Y., Laruelle, G. G., Rödenbeck, C., Roobaert, A., and Zeng, J.: SeaFlux: harmonization of air–sea CO₂ data products using a standardized approach, *Earth System Science Data*, 13, 4693–4710, <https://doi.org/10.5194/essd-13-4693-2021>, 2021.

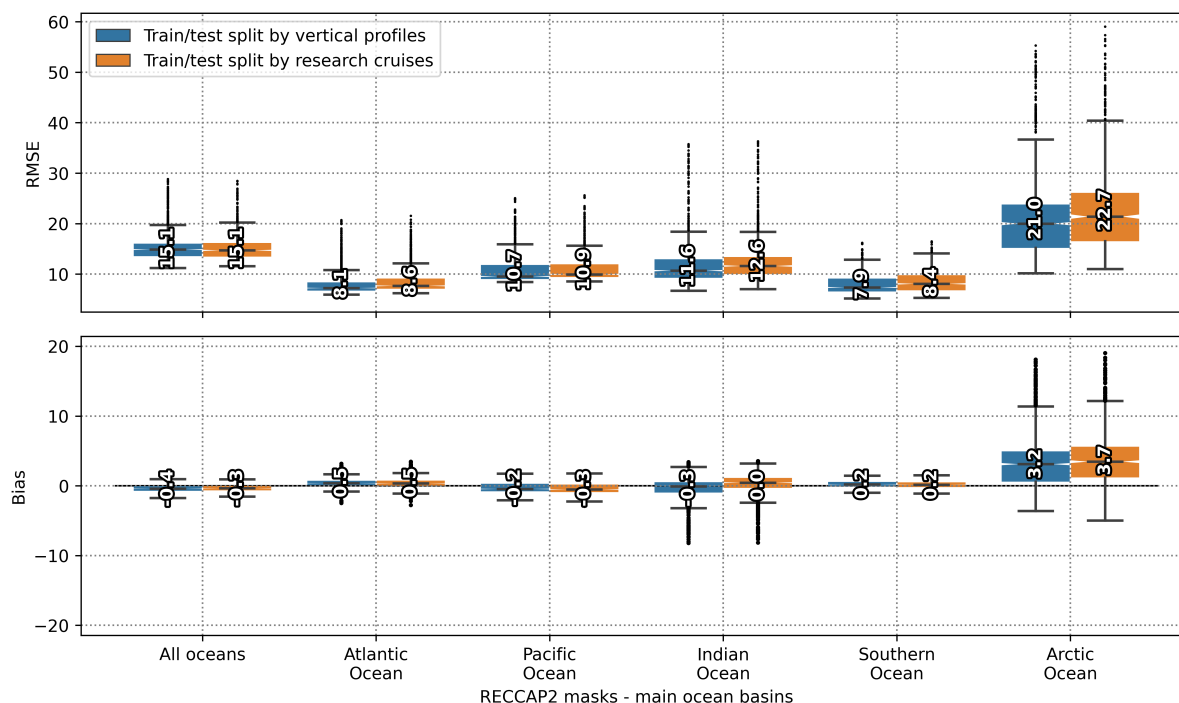


Figure A7. Distribution of RMSE (top panel) and bias (bottom panel) during model validation with ECCO-Darwin (using a model run covering 1992-2017) over different RECCAP2 ocean basins. The blue and orange bars refer to the different error metrics when performing the train/test split that creates the validation dataset: split either by complete vertical profiles (blue) or complete research cruises (orange). Statistics are the results of more than 30 ensemble members in each category. Labels show mean, boxes span the interquartile range, end caps extend to the smallest and largest values within 1.5 times the interquartile range and dots represent outliers. Not shown are the semi-enclosed seas, where the split by research cruises is performing better, leading to similar metrics for all oceans.

Ford, D. J., Blannin, J., Watts, J., Watson, A. J., Landschützer, P., Jersild, A., and Shutler, J. D.: A Comprehensive Analysis of Air-Sea CO₂ Flux Uncertainties Constructed From Surface Ocean Data Products, *Global Biogeochemical Cycles*, 38, <https://doi.org/10.1029/2024gb008188>, 2024.

Friedlingstein, P., O'Sullivan, M., Jones, M. W., Andrew, R. M., Hauck, J., Landschützer, P., Le Quéré, C., Li, H., Luijckx, I. T., Olsen, A., Peters, G. P., Peters, W., Pongratz, J., Schwingshackl, C., Sitch, S., Canadell, J. G., Ciais, P., Jackson, R. B., Alin, S. R., Arneeth, A., Arora, V., Bates, N. R., Becker, M., Bellouin, N., Berghoff, C. F., Bittig, H. C., Bopp, L., Cadule, P., Campbell, K., Chamberlain, M. A., Chandra, N., Chevallier, F., Chini, L. P., Colligan, T., Decayeux, J., Djeutchouang, L. M., Dou, X., Duran Rojas, C., Enyo, K., Evans, W., Fay, A. R., Feely, R. A., Ford, D. J., Foster, A., Gasser, T., Gehlen, M., Gkritzalis, T., Grassi, G., Gregor, L., Gruber, N., Gürses, Ö., Harris, I., Hefner, M., Heinke, J., Hurtt, G. C., Iida, Y., Ilyina, T., Jacobson, A. R., Jain, A. K., Jarnikova, T., Jersild, A., Jiang, F., Jin, Z., Kato, E., Keeling, R. F., Klein Goldewijk, K., Knauer, J., Korsbakken, J. I., Lan, X., Lauvset, S. K., Lefèvre, N., Liu, Z., Liu, J., Ma, L., Maksyutov, S., Marland, G., Mayot, N., McGuire, P. C., Metzl, N., Monacchi, N. M., Morgan, E. J., Nakaoka, S.-I., Neill, C., Niwa, Y., Nützel, T., Olivier, L., Ono, T., Palmer, P. I., Pierrot, D., Qin, Z., Resplandy, L., Roobaert, A., Rosan, T. M., Rödenbeck, C., Schwinger,

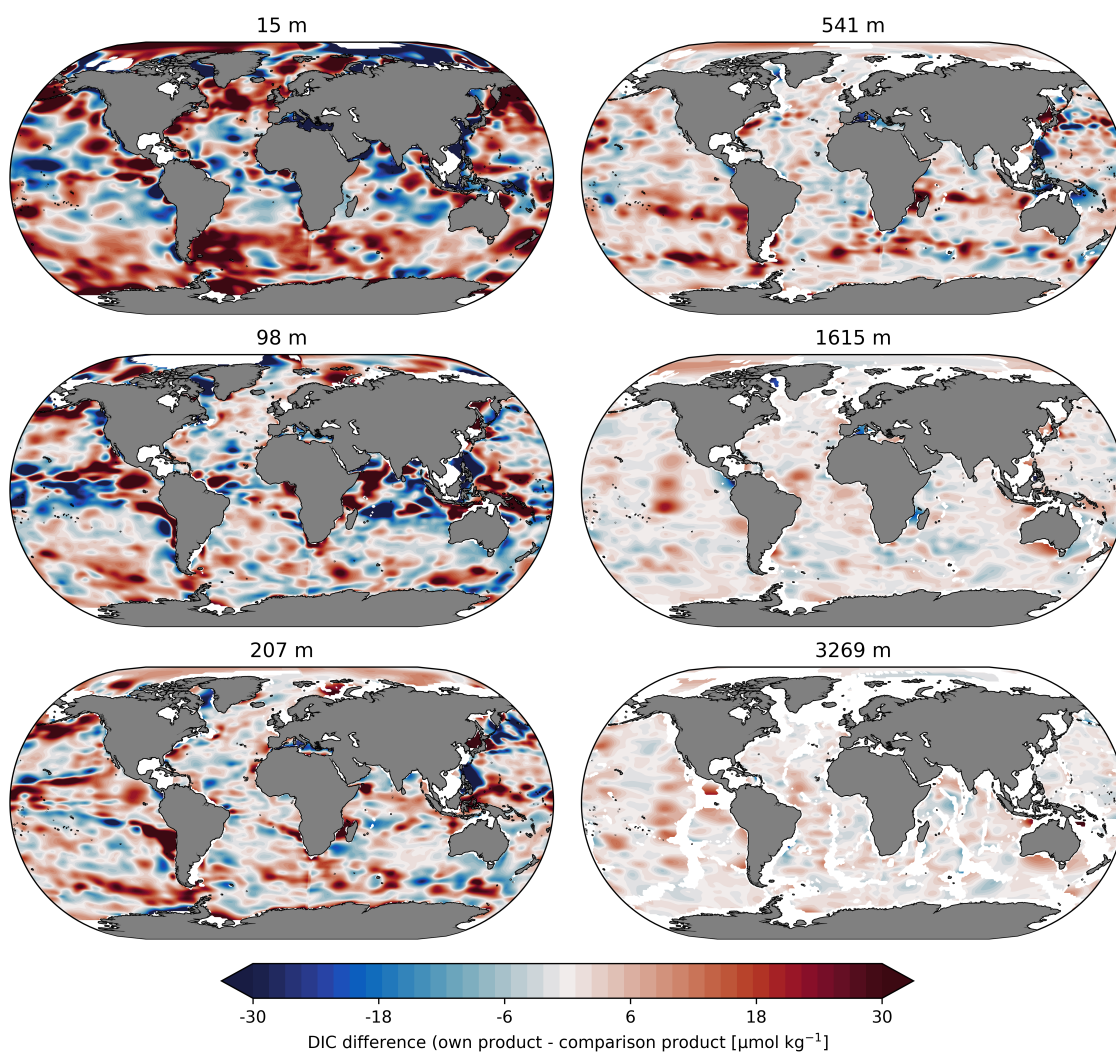


Figure A8. Comparison of DIC values with GLODAPv2.2016b climatology (Lauvset et al. 2016). Prior to taking the difference this product is interpolated to the spatial grid of the GLODAPv2.2016b climatology. As DIC is normalised to the year 2002 in GLODAPv2.2016b, the mean of 2002 is also taken for this product to compare the two.

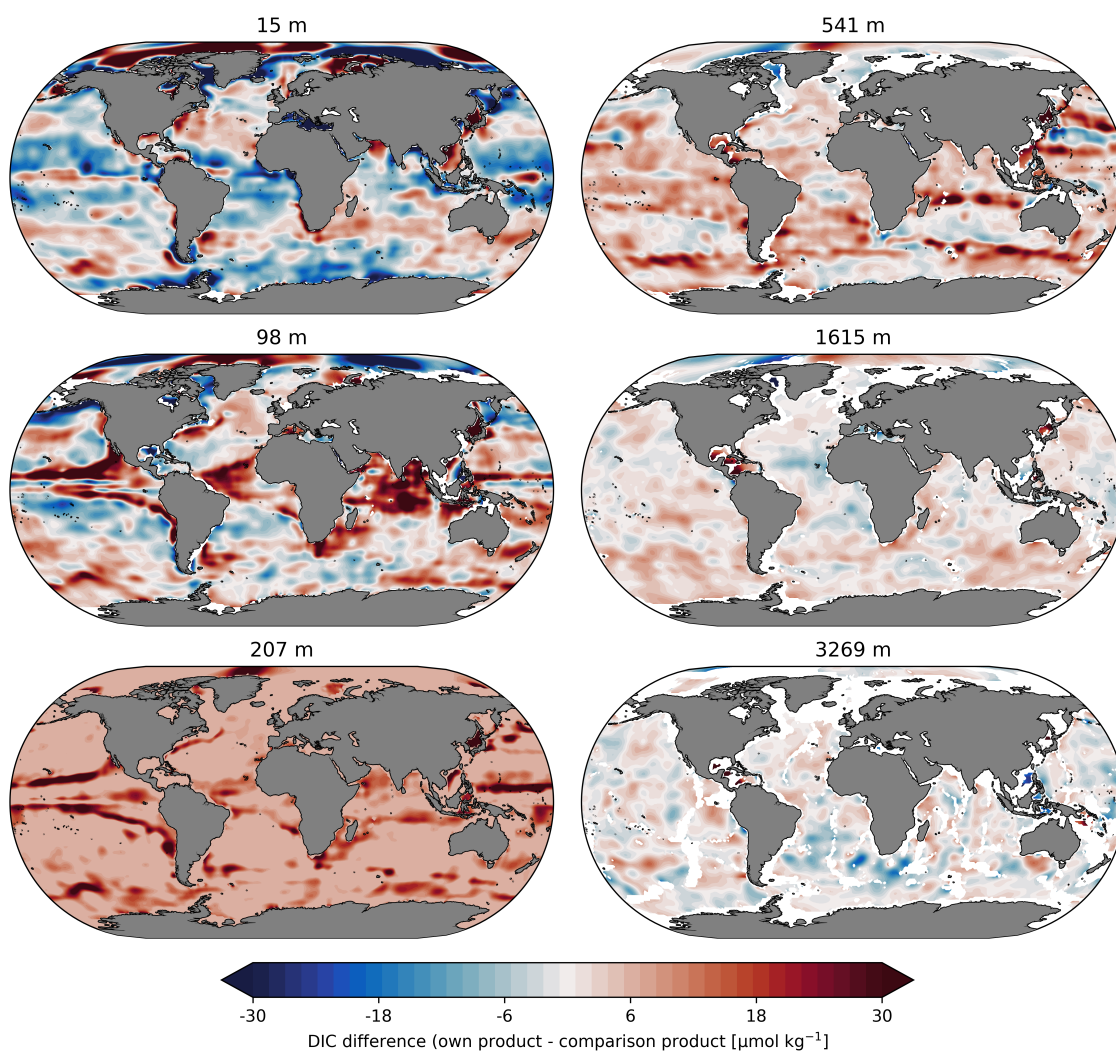


Figure A9. Comparison of DIC values with the NNGv2LDEO climatology (Broullón et al., 2020). Prior to taking the difference this product is interpolated to the spatial grid of the NNGv2LDEO climatology for the year 1995.

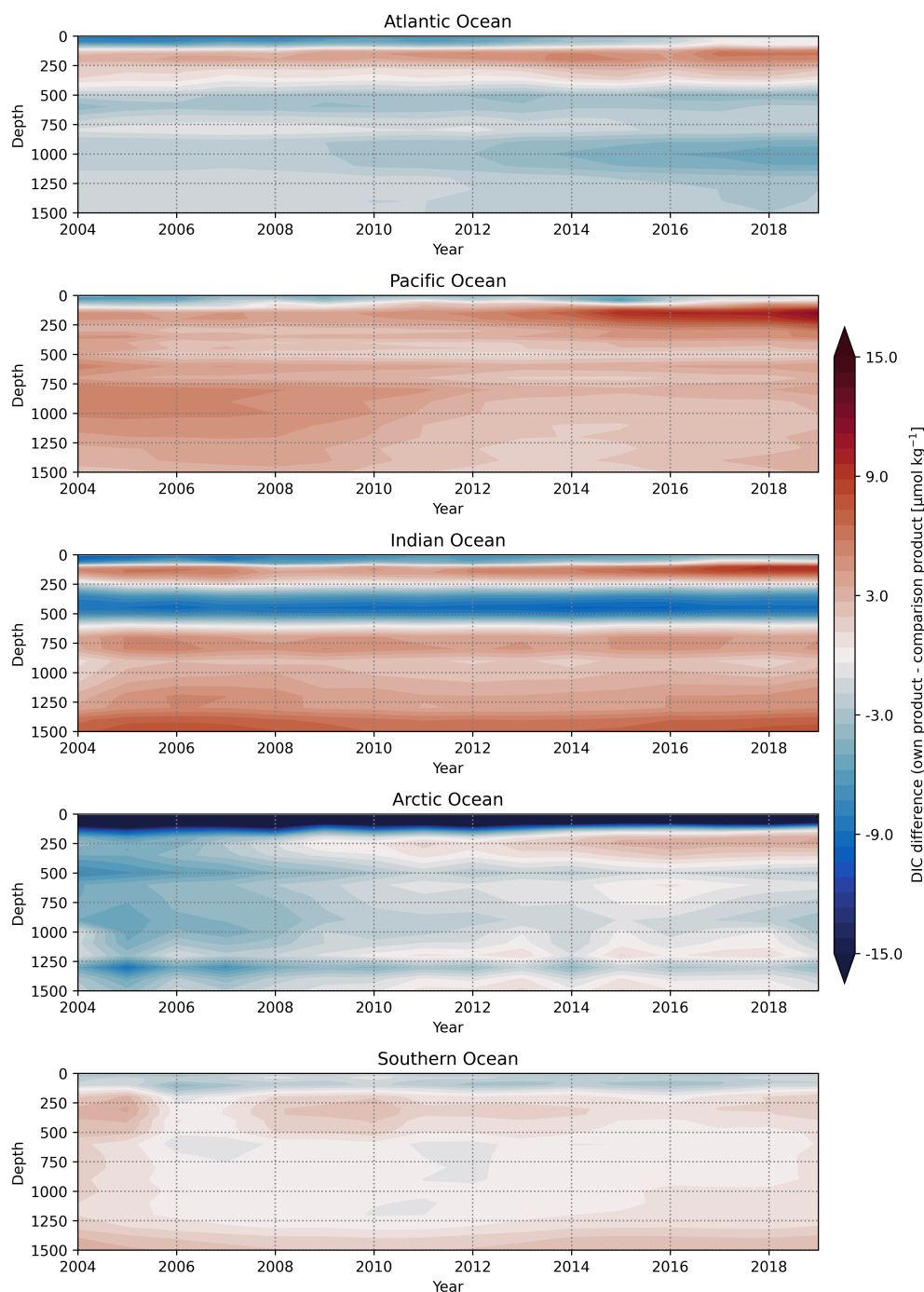


Figure A10. Hovmöller plots showing the difference between our DIC reconstruction and the MOBO-DIC reconstruction (Keppler et al., 2023) over time.

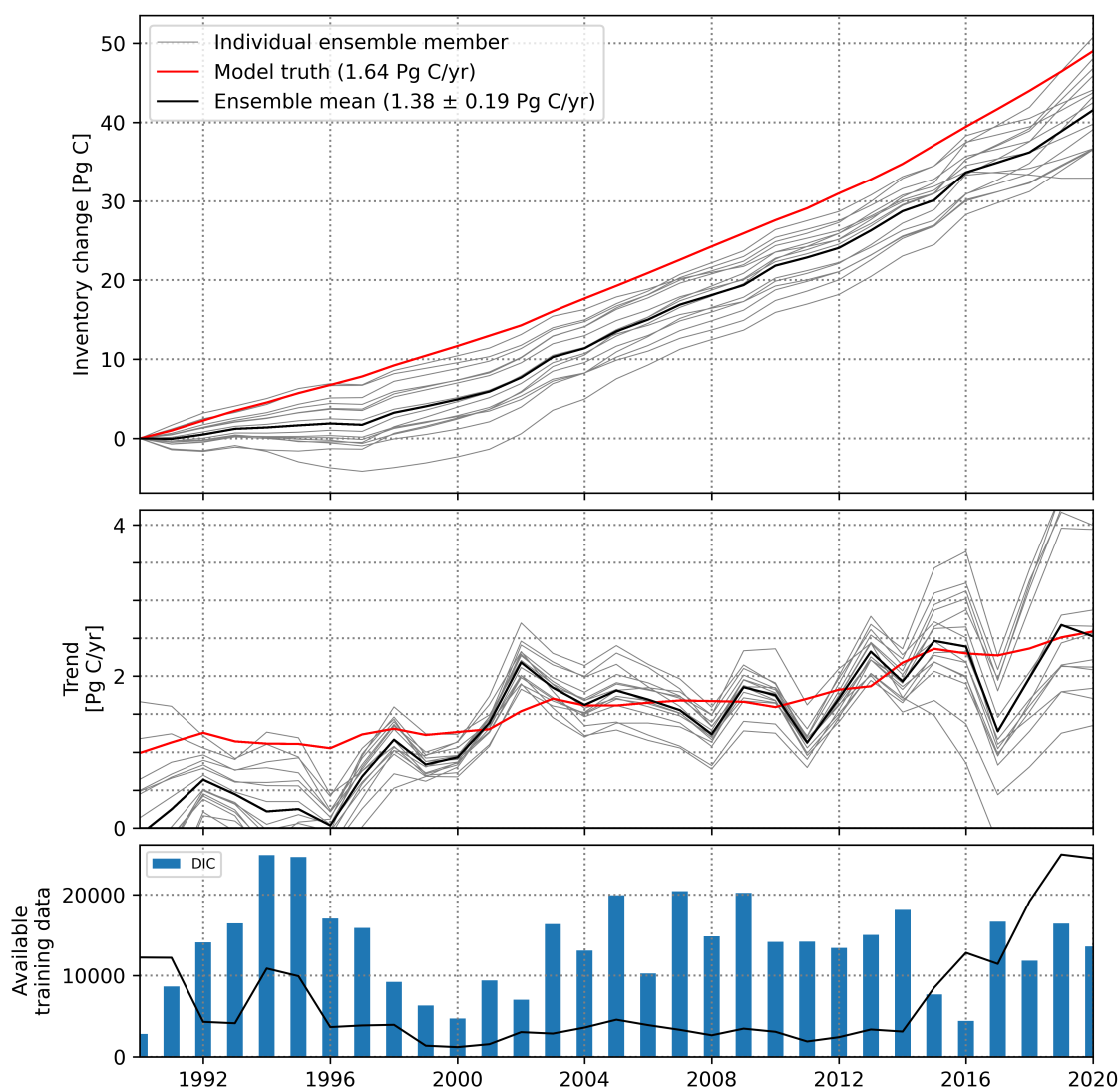


Figure A11. Annual mean of the carbon inventory change since 1990 (top panel) and uptake trend (middle panel, positive values indicating uptake of carbon) based on the model validation through training on a subset of ECCO-Darwin, followed by a prediction of the full model fields. Model truth shown in red, individual ensemble members in grey, ensemble mean in black. Bottom panel shows the available training data compared to a line showing changing standard deviation within the ensemble.

665 J., Smallman, T. L., Smith, S. M., Sospedra-Alfonso, R., Steinhoff, T., Sun, Q., Sutton, A. J., Séférian, R., Takao, S., Tatebe, H., Tian, H., Tilbrook, B., Torres, O., Tourigny, E., Tsujino, H., Tubiello, F., van der Werf, G., Wanninkhof, R., Wang, X., Yang, D., Yang, X., Yu, Z., Yuan, W., Yue, X., Zaehle, S., Zeng, N., and Zeng, J.: Global Carbon Budget 2024, *Earth System Science Data*, 17, 965–1039, <https://doi.org/10.5194/essd-17-965-2025>, 2024.

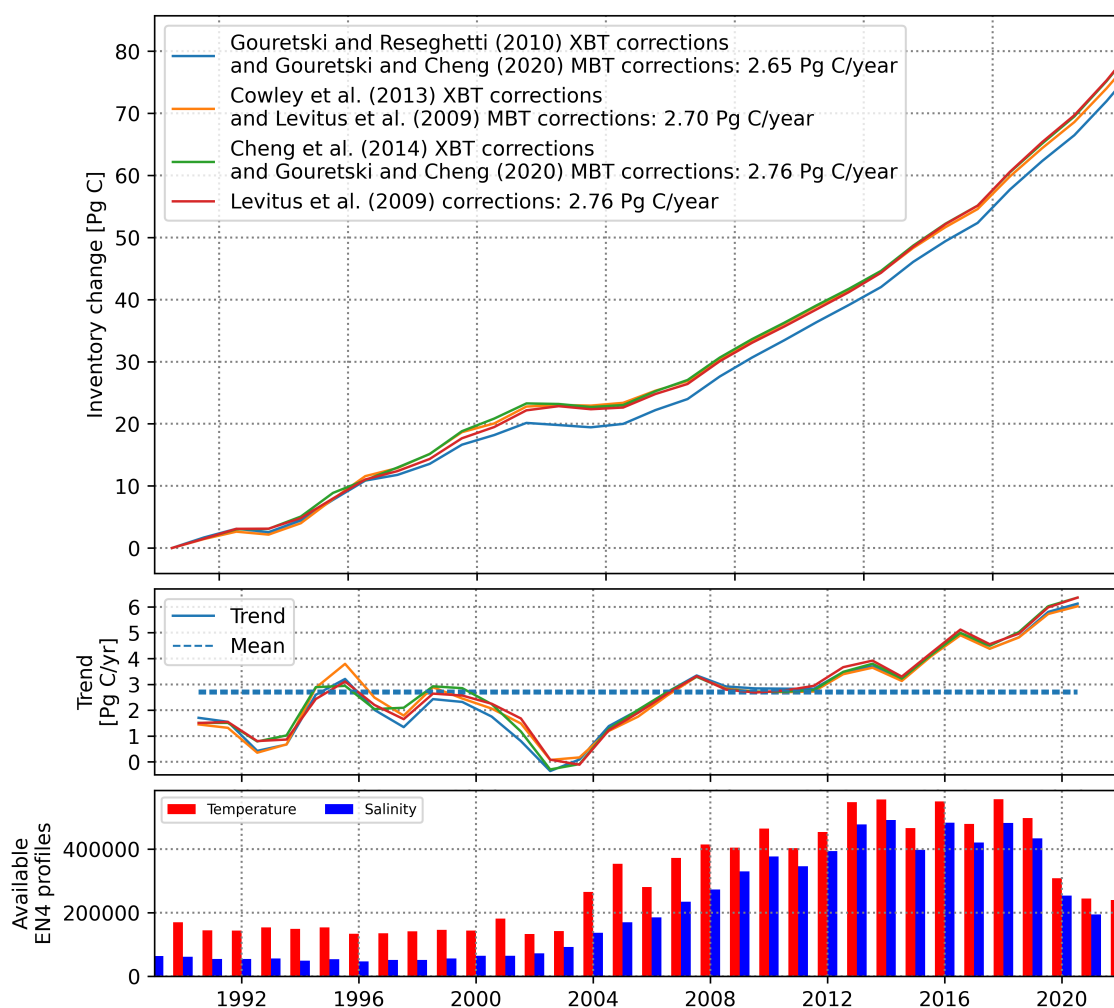


Figure A12. Carbon inventory change from four machine learning ensemble means, each mapped using distinct temperature and salinity corrections applied to the EN4 analysis dataset.

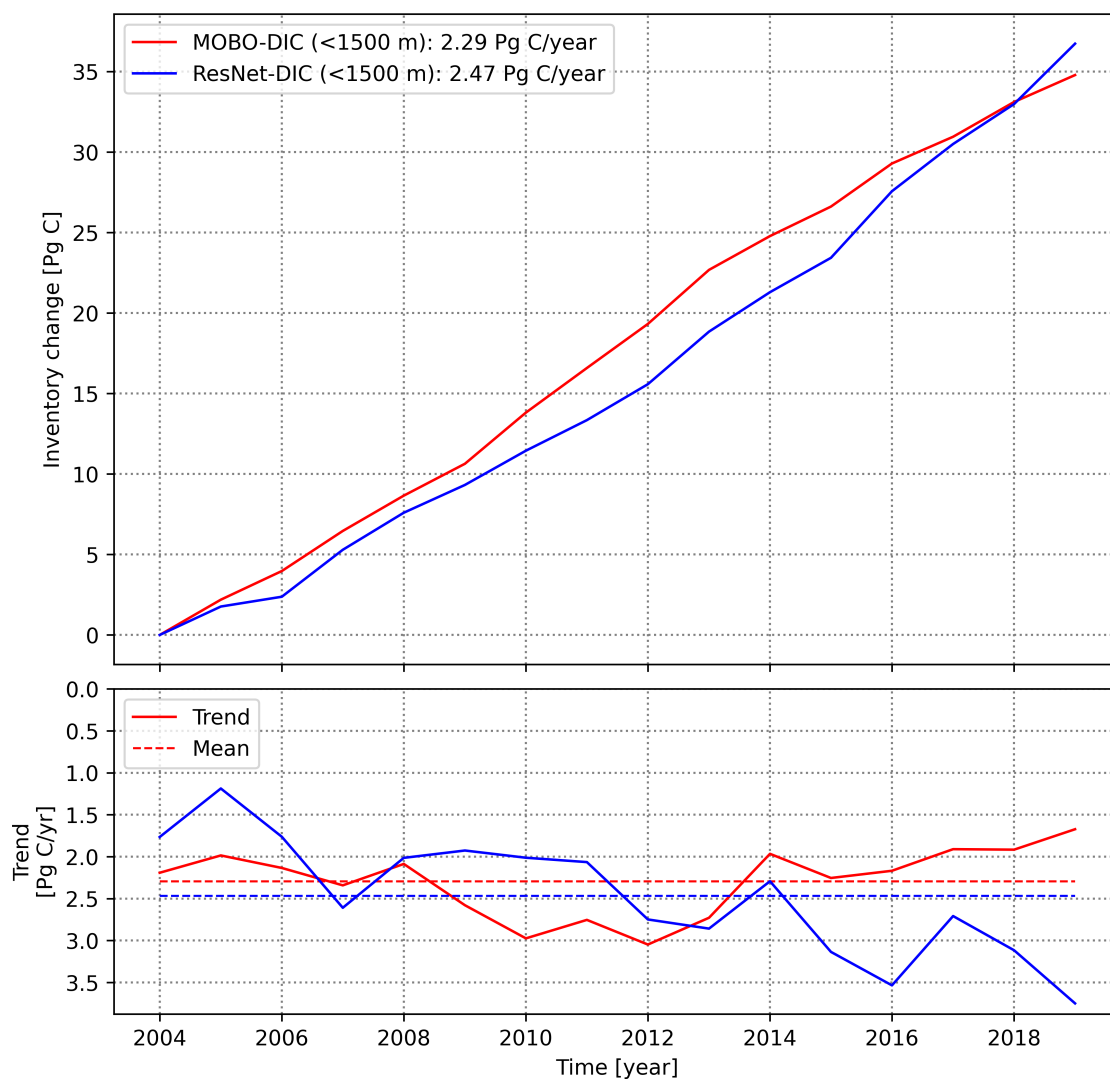


Figure A13. Comparison of carbon inventory change (top panel) and uptake trend (bottom panel, positive values indicating uptake of carbon) between MOBO-DIC (Keppler et al., 2023) (red) and ResNet-DIC (blue). Carbon inventory change is relative to 2004, annualised for MOBO-DIC and ResNet-DIC and restricted to 1500 m, the maximum depth of MOBO-DIC.

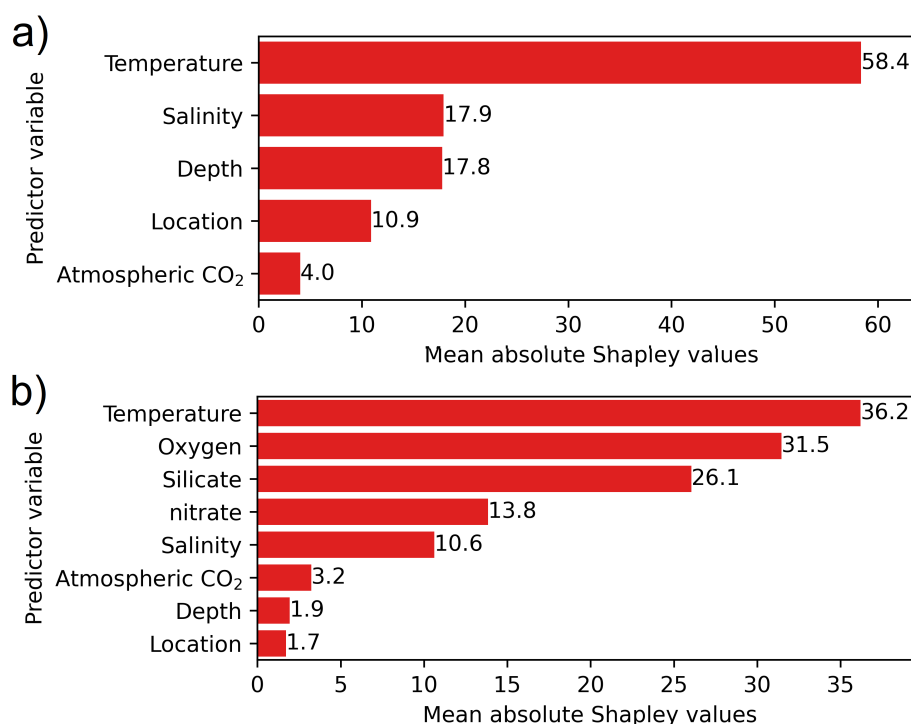


Figure A14. Shapley values of different sets predictor variables, showing the average impact on the DIC reconstruction. Predictor variables used in our reconstructions are shown in a); an alternative scenario with added oxygen, nitrate and silicate is shown in b). The latter case is based on separate training without the possibility of mapping it on a global grid. Note that temperature still has the highest impact on the reconstruction, but oxygen, silicate and nitrate are more important than salinity.

Friedlingstein, P., O'Sullivan, M., Jones, M. W., Andrew, R. M., Bakker, D. C. E., Hauck, J., Landschützer, P., Le Quéré, C., Li, H., Luijkx,
 670 I. T., Peters, G. P., Peters, W., Pongratz, J., Schwingshackl, C., Sitch, S., Canadell, J. G., Ciais, P., Aas, K., Alin, S. R., Anthoni, P.,
 Barbero, L., Bates, N. R., Bellouin, N., Benoit-Cattin, A., Berghoff, C. F., Bernardello, R., Bopp, L., Brasika, I. B. M., Chamberlain,
 M. A., Chandra, N., Chevallier, F., Chini, L. P., Collier, N. O., Colligan, T. H., Cronin, M., Djeutchouang, L. M., Dou, X., Enright, M. P.,
 Enyo, K., Erb, M., Evans, W., Feely, R. A., Feng, L., Ford, D. J., Foster, A., Fransner, F., Gasser, T., Gehlen, M., Gkritzalis, T., Goncalves
 De Souza, J., Grassi, G., Gregor, L., Gruber, N., Guenet, B., Gürses, Ö., Harrington, K., Harris, I., Heinke, J., Hurtt, G. C., Iida, Y., Ilyina,
 675 T., Ito, A., Jacobson, A. R., Jain, A. K., Jarníková, T., Jersild, A., Jiang, F., Jones, S. D., Kato, E., Keeling, R. F., Klein Goldewijk, K.,
 Knauer, J., Kong, Y., Korsbakken, J. I., Koven, C., Kunimitsu, T., Lan, X., Liu, J., Liu, Z., Liu, Z., Lo Monaco, C., Ma, L., Marland,
 G., McGuire, P. C., McKinley, G. A., Melton, J. R., Monacchi, N., Monier, E., Morgan, E. J., Munro, D. R., Müller, J. D., Nakaoka, S.-I.,
 Nayagam, L. R., Niwa, Y., Nutzelt, T., Olsen, A., Omar, A. M., Pan, N., Pandey, S., Pierrot, D., Qin, Z., Regnier, P., Rehder, G., Resplandy,
 L., Roobaert, A., Rosan, T. M., Rödenbeck, C., Schwinger, J., Skjelvan, I., Smallman, T. L., Spada, V., Sreeush, M. G., Sun, Q., Sutton,
 680 A. J., Sweeney, C., Swingedouw, D., Séférian, R., Takao, S., Tatebe, H., Tian, H., Tian, X., Tilbrook, B., Tsujino, H., Tubiello, F., van
 Ooijen, E., van der Werf, G. R., van de Velde, S. J., Walker, A. P., Wanninkhof, R., Yang, X., Yuan, W., Yue, X., and Zeng, J.: Global
 Carbon Budget 2025, *Earth System Science Data*, 18, 3211–3288, <https://doi.org/10.5194/essd-18-3211-2026>, 2026.

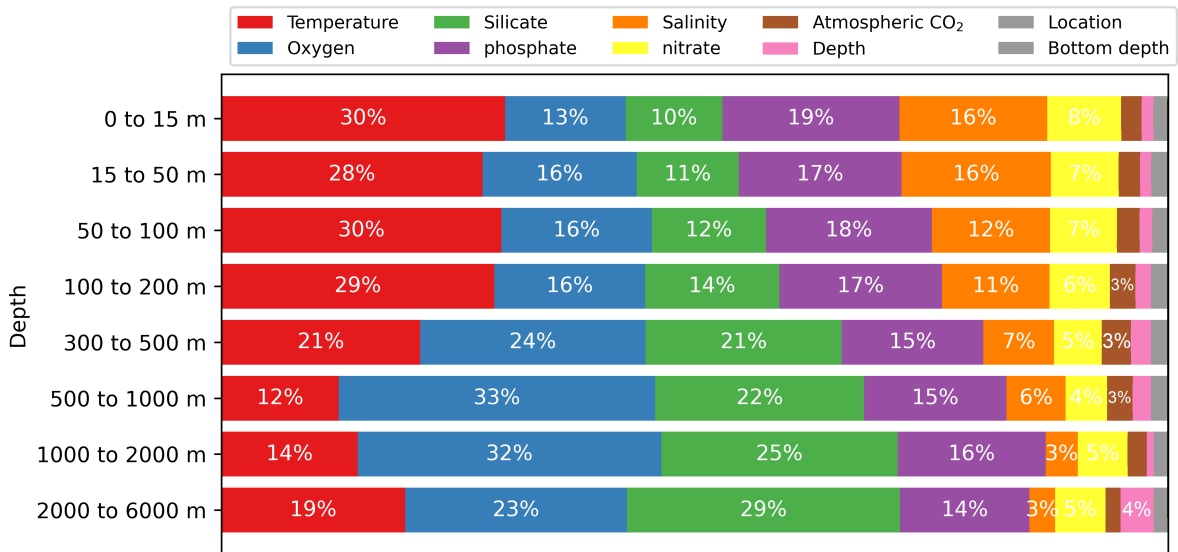


Figure A15. Shapley values of different predictor variables and their relative importance with increasing depth.

Fröb, F., Olsen, A., Pérez, F. F., García-Ibáñez, M. I., Jeansson, E., Omar, A., and Lauvset, S. K.: Inorganic carbon and water masses in the Irminger Sea since 1991, *Biogeosciences*, 15, 51–72, <https://doi.org/10.5194/bg-15-51-2018>, 2018.

685 Gade, K.: A Non-singular Horizontal Position Representation, *Journal of Navigation*, 63, 395–417, <https://doi.org/10.1017/s0373463309990415>, 2010.

Garcia, H., Weathers, K., Paver, C., Smolyar, I., Boyer, T., Locarnini, M., Zweng, M., Mishonov, A., Baranova, O., Seidov, D., and Reagan, J.: World Ocean Atlas 2018, Volume 3: Dissolved Oxygen, Apparent Oxygen Utilization, and Dissolved Oxygen Saturation., Report, <https://archimer.ifremer.fr/doc/00651/76337/>, 2019.

690 GEBCO Bathymetric Compilation Group 2024: The GEBCO 2024 Grid - a continuous terrain model of the global oceans and land., <https://doi.org/10.5285/1C44CE99-0A0D-5F4F-E063-7086ABC0EA0F>, 2024.

Gloege, L., McKinley, G. A., Landschützer, P., Fay, A. R., Frölicher, T. L., Fyfe, J. C., Ilyina, T., Jones, S., Lovenduski, N. S., Rodgers, K. B., Schlunegger, S., and Takano, Y.: Quantifying Errors in Observationally Based Estimates of Ocean Carbon Sink Variability, *Global Biogeochemical Cycles*, 35, <https://doi.org/10.1029/2020gb006788>, 2021.

695 Good, S. A., Martin, M. J., and Rayner, N. A.: EN4: Quality controlled ocean temperature and salinity profiles and monthly objective analyses with uncertainty estimates, *Journal of Geophysical Research: Oceans*, 118, 6704–6716, <https://doi.org/10.1002/2013jc009067>, 2013.

Gregor, L. and Gruber, N.: OceanSODA-ETHZ: a global gridded data set of the surface ocean carbonate system for seasonal to decadal studies of ocean acidification, *Earth System Science Data*, 13, 777–808, <https://doi.org/10.5194/essd-13-777-2021>, 2021.

Gruber, N. and Sarmiento, J. L.: Chapter 9. Large-scale Biogeochemical–physical Interactions in Elemental Cycles, *The Sea*, 12, 337–399, <https://www.semanticscholar.org/paper/9c463346fa2d3363208df4046e4920ce8b71c187>, 2002.

700



- Gruber, N., Clement, D., Carter, B. R., Feely, R. A., van Heuven, S., Hoppema, M., Ishii, M., Key, R. M., Kozyr, A., Lauvset, S. K., Monaco, C. L., Mathis, J. T., Murata, A., Olsen, A., Perez, F. F., Sabine, C. L., Tanhua, T., and Wanninkhof, R.: The oceanic sink for anthropogenic CO₂ from 1994 to 2007, *Science*, 363, 1193–1199, <https://doi.org/10.1126/science.aau5153>, 2019.
- Gruber, N., Bakker, D. C. E., DeVries, T., Gregor, L., Hauck, J., Landschützer, P., McKinley, G. A., and Müller, J. D.: Trends and variability
 705 in the ocean carbon sink, *Nature Reviews Earth and Environment*, <https://doi.org/10.1038/s43017-022-00381-x>, 2023.
- Grégoire, M., Garçon, V., Garcia, H., Breitburg, D., Isensee, K., Oschlies, A., Telszewski, M., Barth, A., Bittig, H. C., Carstensen, J., Carval, T., Chai, F., Chavez, F., Conley, D., Coppola, L., Crowe, S., Currie, K., Dai, M., Deflandre, B., Dewitte, B., Diaz, R., Garcia-Robledo, E., Gilbert, D., Giorgetti, A., Glud, R., Gutierrez, D., Hosoda, S., Ishii, M., Jacinto, G., Langdon, C., Lauvset, S. K., Levin, L. A., Limburg, K. E., Mehrtens, H., Montes, I., Naqvi, W., Paulmier, A., Pfeil, B., Pitcher, G., Pouliquen, S., Rabalais, N., Rabouille, C., Recape, V., Roman, M., Rose, K., Rudnick, D., Rummer, J., Schmechtig, C., Schmidtke, S., Seibel, B., Slomp, C., Sumalia, U. R., Tanhua, T., Thierry, V.,
 710 Uchida, H., Wanninkhof, R., and Yasuhara, M.: A Global Ocean Oxygen Database and Atlas for Assessing and Predicting Deoxygenation and Ocean Health in the Open and Coastal Ocean, *Frontiers in Marine Science*, 8, <https://doi.org/10.3389/fmars.2021.724913>, 2021.
- Hauck, J., Zeising, M., Quéré, C. L., Gruber, N., Bakker, D. C. E., Bopp, L., Chau, T. T. T., Özgür Gürses, Ilyina, T., Landschützer, P., Lenton, A., Resplandy, L., Rödenbeck, C., Schwinger, J., and Séférian, R.: Consistency and Challenges in the Ocean Carbon Sink Estimate for the
 715 Global Carbon Budget, *Frontiers in Marine Science*, 7, <https://doi.org/10.3389/fmars.2020.571720>, 2020.
- Hauck, J., Nissen, C., Landschützer, P., Rödenbeck, C., Bushinsky, S., and Olsen, A.: Sparse observations induce large biases in estimates of the global ocean CO₂ sink: an ocean model subsampling experiment, *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 381, <https://doi.org/10.1098/rsta.2022.0063>, 2023.
- He, K., Zhang, X., Ren, S., and Sun, J.: Deep Residual Learning for Image Recognition, <https://doi.org/10.48550/ARXIV.1512.03385>, 2015.
- Hennige, S. J., Wolfram, U., Wickes, L., Murray, F., Roberts, J. M., Kamenos, N. A., Schofield, S., Groetsch, A., Spiesz, E. M., Aubin-Tam, M.-E., and Etnoyer, P. J.: Crumbling Reefs and Cold-Water Coral Habitat Loss in a Future Ocean: Evidence of “Coralporosis” as an
 720 Indicator of Habitat Integrity, *Frontiers in Marine Science*, 7, <https://doi.org/10.3389/fmars.2020.00668>, 2020.
- Intergovernmental Panel on Climate Change: Climate Change 2021 – The Physical Science Basis: Working Group I Contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, Cambridge University Press, ISBN 9781009157896,
 725 <https://doi.org/10.1017/9781009157896>, 2023.
- Ito, T., Cervania, A., Cross, K., Ainchwar, S., and Delawalla, S.: Mapping Dissolved Oxygen Concentrations by Combining Shipboard and Argo Observations Using Machine Learning Algorithms, *Journal of Geophysical Research: Machine Learning and Computation*, 1, <https://doi.org/10.1029/2024jh000272>, 2024.
- Keppler, L. and Landschützer, P.: Regional Wind Variability Modulates the Southern Ocean Carbon Sink, *Scientific Reports*, 9,
 730 <https://doi.org/10.1038/s41598-019-43826-y>, 2019.
- Keppler, L., Landschützer, P., Gruber, N., Lauvset, S. K., and Stemmler, I.: Seasonal Carbon Dynamics in the Near-Global Ocean, *Global Biogeochemical Cycles*, 34, <https://doi.org/10.1029/2020gb006571>, 2020.
- Keppler, L., Landschützer, P., Lauvset, S. K., and Gruber, N.: Recent Trends and Variability in the Oceanic Storage of Dissolved Inorganic Carbon, *Global Biogeochemical Cycles*, 37, <https://doi.org/10.1029/2022gb007677>, 2023.
- Kohonen, T.: Self-Organizing Feature Maps, pp. 119–157, Springer Berlin Heidelberg, ISBN 9783642881633, https://doi.org/10.1007/978-3-642-88163-3_5, 1989.
- Kohonen, T.: Self-Organizing Maps, Springer Berlin Heidelberg, ISBN 9783642569272, <https://doi.org/10.1007/978-3-642-56927-2>, 2001.



- Kwiatkowski, L., Torres, O., Bopp, L., Aumont, O., Chamberlain, M., Christian, J. R., Dunne, J. P., Gehlen, M., Ilyina, T., John, J. G., Lenton, A., Li, H., Lovenduski, N. S., Orr, J. C., Palmieri, J., Santana-Falcón, Y., Schwinger, J., Séférian, R., Stock, C. A., Tagliabue, A., Takano, Y., Tjiputra, J., Toyama, K., Tsujino, H., Watanabe, M., Yamamoto, A., Yool, A., and Ziehn, T.: Twenty-first century ocean warming, acidification, deoxygenation, and upper-ocean nutrient and primary production decline from CMIP6 model projections, *Biogeosciences*, 17, 3439–3470, <https://doi.org/10.5194/bg-17-3439-2020>, 2020.
- 740 Landschützer, P., Gruber, N., Bakker, D. C. E., Schuster, U., Nakaoka, S., Payne, M. R., Sasse, T. P., and Zeng, J.: A neural network-based estimate of the seasonal to inter-annual variability of the Atlantic Ocean carbon sink, *Biogeosciences*, 10, 7793–7815, <https://doi.org/10.5194/bg-10-7793-2013>, 2013.
- 745 Landschützer, P., Gruber, N., Bakker, D. C. E., and Schuster, U.: Recent variability of the global ocean carbon sink, *Global Biogeochemical Cycles*, 28, 927–949, <https://doi.org/10.1002/2014gb004853>, 2014.
- Landschützer, P., Gruber, N., Haumann, F. A., Rödenbeck, C., Bakker, D. C. E., van Heuven, S., Hoppema, M., Metzl, N., Sweeney, C., Takahashi, T., Tilbrook, B., and Wanninkhof, R.: The reinvigoration of the Southern Ocean carbon sink, *Science*, 349, 1221–1224, <https://doi.org/10.1126/science.aab2620>, 2015.
- 750 Landschützer, P., Gruber, N., and Bakker, D. C. E.: Decadal variations and trends of the global ocean carbon sink, *Global Biogeochemical Cycles*, 30, 1396–1417, <https://doi.org/10.1002/2015gb005359>, 2016.
- Landschützer, P., Laruelle, G. G., Roobaert, A., and Regnier, P.: A uniform pCO₂ climatology combining open and coastal oceans, *Earth System Science Data*, 12, 2537–2553, <https://doi.org/10.5194/essd-12-2537-2020>, 2020.
- 755 Laruelle, G. G., Landschützer, P., Gruber, N., Tison, J.-L., Delille, B., and Regnier, P.: Global high-resolution monthly pCO₂ climatology for the coastal ocean derived from neural network interpolation, *Biogeosciences*, 14, 4545–4561, <https://doi.org/10.5194/bg-14-4545-2017>, 2017.
- Lauvset, S. K., Key, R. M., Olsen, A., van Heuven, S., Velo, A., Lin, X., Schirnick, C., Kozyr, A., Tanhua, T., Hoppema, M., Jutterström, S., Steinfeldt, R., Jeansson, E., Ishii, M., Perez, F. F., Suzuki, T., and Watelet, S.: A new global interior ocean mapped climatology: the 1×1 degree GLODAP version 2, *Earth System Science Data*, 8, 325–340, <https://doi.org/10.5194/essd-8-325-2016>, 2016.
- 760 Lauvset, S. K., Lange, N., Tanhua, T., Bittig, H. C., Olsen, A., Kozyr, A., Álvarez, M., Becker, S., Brown, P. J., Carter, B. R., Cotrim da Cunha, L., Feely, R. A., van Heuven, S., Hoppema, M., Ishii, M., Jeansson, E., Jutterström, S., Jones, S. D., Karlsen, M. K., Lo Monaco, C., Michaelis, P., Murata, A., Pérez, F. F., Pfeil, B., Schirnick, C., Steinfeldt, R., Suzuki, T., Tilbrook, B., Velo, A., Wanninkhof, R., Woosley, R. J., and Key, R. M.: An updated version of the global interior ocean biogeochemical data product, GLODAPv2.2021, *Earth System Science Data*, 13, 5565–5589, <https://doi.org/10.5194/essd-13-5565-2021>, 2021.
- 765 Lauvset, S. K., Lange, N., Tanhua, T., Bittig, H. C., Olsen, A., Kozyr, A., Álvarez, M., Azetsu-Scott, K., Brown, P. J., Carter, B. R., Cotrim da Cunha, L., Hoppema, M., Humphreys, M. P., Ishii, M., Jeansson, E., Murata, A., Müller, J. D., Perez, F. F., Schirnick, C., Steinfeldt, R., Suzuki, T., Ulfsbo, A., Velo, A., Woosley, R. J., and Key, R.: The annual update GLODAPv2.2023: the global interior ocean biogeochemical data product, *Earth System Science Data*, <https://doi.org/10.5194/essd-2023-468>, 2024.
- 770 Le Quéré, C., Rödenbeck, C., Buitenhuis, E. T., Conway, T. J., Langenfelds, R., Gomez, A., Labuschagne, C., Ramonet, M., Nakazawa, T., Metzl, N., Gillett, N., and Heimann, M.: Saturation of the Southern Ocean CO₂ Sink Due to Recent Climate Change, *Science*, 316, 1735–1738, <https://doi.org/10.1126/science.1136188>, 2007.
- Lundberg, S. M. and Lee, S.-I.: A Unified Approach to Interpreting Model Predictions, in: *Advances in Neural Information Processing Systems*, edited by Guyon, I., Luxburg, U. V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., and Garnett, R., vol. 30, Curran Associates, Inc., https://proceedings.neurips.cc/paper_files/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf, 2017.
- 775



- Matear, R. J. and Hirst, A. C.: Climate change feedback on the future oceanic CO₂ uptake, *Tellus B: Chemical and Physical Meteorology*, 51, 722, <https://doi.org/10.3402/tellusb.v51i3.16472>, 1999.
- McKinley, G. A., Fay, A. R., Eddebbar, Y. A., Gloege, L., and Lovenduski, N. S.: External Forcing Explains Recent Decadal Variability of the Ocean Carbon Sink, *AGU Advances*, 1, <https://doi.org/10.1029/2019av000149>, 2020.
- 780 Metzl, N., Fin, J., Lo Monaco, C., Mignon, C., Alliouane, S., Antoine, D., Bourdin, G., Boutin, J., Bozec, Y., Conan, P., Coppola, L., Diaz, F., Douville, E., Durrieu de Madron, X., Gattuso, J.-P., Gazeau, F., Golbol, M., Lansard, B., Lefèvre, D., Lefèvre, N., Lombard, F., Louanchi, F., Merlivat, L., Olivier, L., Petrenko, A., Petton, S., Pujo-Pay, M., Rabouille, C., Reverdin, G., Ridame, C., Tribollet, A., Vellucci, V., Wagener, T., and Wimart-Rousseau, C.: A synthesis of ocean total alkalinity and dissolved inorganic carbon measurements from 1993 to 2022: the SNAPO-CO₂-v1 dataset, *Earth System Science Data*, 16, 89–120, <https://doi.org/10.5194/essd-16-89-2024>, 2024.
- 785 Müller, J. D. and Gruber, N.: Progression of Ocean Interior Acidification over the Industrial Era, *ESS Open Archive*, <https://doi.org/10.22541/essoar.170654370.07309198/v1>, 2024.
- Olsen, A., Key, R. M., van Heuven, S., Lauvset, S. K., Velo, A., Lin, X., Schirnack, C., Kozyr, A., Tanhua, T., Hoppema, M., Jutterström, S., Steinfeldt, R., Jeansson, E., Ishii, M., Pérez, F. F., and Suzuki, T.: The Global Ocean Data Analysis Project version 2 (GLODAPv2) – an internally consistent data product for the world ocean, *Earth System Science Data*, 8, 297–323, <https://doi.org/10.5194/essd-8-297-2016>, 2016.
- 790 Orr, J. C., Fabry, V. J., Aumont, O., Bopp, L., Doney, S. C., Feely, R. A., Gnanadesikan, A., Gruber, N., Ishida, A., Joos, F., Key, R. M., Lindsay, K., Maier-Reimer, E., Matear, R., Monfray, P., Mouchet, A., Najjar, R. G., Plattner, G.-K., Rodgers, K. B., Sabine, C. L., Sarmiento, J. L., Schlitzer, R., Slater, R. D., Totterdell, I. J., Weirig, M.-F., Yamanaka, Y., and Yool, A.: Anthropogenic ocean acidification over the twenty-first century and its impact on calcifying organisms, *Nature*, 437, 681–686, <https://doi.org/10.1038/nature04095>, 2005.
- 795 Raimondi, L., Tanhua, T., Azetsu-Scott, K., Yashayaev, I., and Wallace, D.: A 30 -Year Time Series of Transient Tracer-Based Estimates of Anthropogenic Carbon in the Central Labrador Sea, *Journal of Geophysical Research: Oceans*, 126, <https://doi.org/10.1029/2020jc017092>, 2021.
- Regnier, P., Resplandy, L., Najjar, R. G., and Ciais, P.: The land-to-ocean loops of the global carbon cycle, *Nature*, 603, 401–410, <https://doi.org/10.1038/s41586-021-04339-9>, 2022.
- 800 Sarmiento, J. L. and Gruber, N.: *Ocean Biogeochemical Dynamics*, Princeton University Press, ISBN 9780691017075, <https://doi.org/10.2307/j.ctt3fgxqx>, 2013.
- Sarmiento, J. L. and Sundquist, E. T.: Revised budget for the oceanic uptake of anthropogenic carbon dioxide, *Nature*, 356, 589–593, <https://doi.org/10.1038/356589a0>, 1992.
- Sarmiento, J. L., Orr, J. C., and Siegenthaler, U.: A perturbation simulation of CO₂ uptake in an ocean general circulation model, *Journal of Geophysical Research: Oceans*, 97, 3621–3645, <https://doi.org/10.1029/91jc02849>, 1992.
- 805 Seibel, B. A. and Walsh, P. J.: Potential Impacts of CO₂ Injection on Deep-Sea Biota, *Science*, 294, 319–320, <https://doi.org/10.1126/science.1065301>, 2001.
- Seland, Ø., Bentsen, M., Olivé, D., Toniazzo, T., Gjermundsen, A., Graff, L. S., Debernard, J. B., Gupta, A. K., He, Y.-C., Kirkevåg, A., Schwinger, J., Tjiputra, J., Aas, K. S., Bethke, I., Fan, Y., Griesfeller, J., Grini, A., Guo, C., Ilıcak, M., Karset, I. H. H., Landgren, O., Liakka, J., Moseid, K. O., Nummelin, A., Spensberger, C., Tang, H., Zhang, Z., Heinze, C., Iversen, T., and Schulz, M.: Overview of the Norwegian Earth System Model (NorESM2) and key climate response of CMIP6 DECK, historical, and scenario simulations, *Geoscientific Model Development*, 13, 6165–6200, <https://doi.org/10.5194/gmd-13-6165-2020>, 2020.
- Shapley, L. S.: Notes on the n-person game II: The value of an n-person game, Research memorandum, 1951.



- Sharp, J. D., Fassbender, A. J., Carter, B. R., Johnson, G. C., Schultz, C., and Dunne, J. P.: GOBAI-O2: temporally and spatially resolved
 815 fields of ocean interior dissolved oxygen over nearly 2 decades, *Earth System Science Data*, 15, 4481–4518, <https://doi.org/10.5194/essd-15-4481-2023>, 2023.
- Slack, D., Hilgard, S., Jia, E., Singh, S., and Lakkaraju, H.: Fooling LIME and SHAP: Adversarial Attacks on Post hoc Explanation Methods, in: *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society, AIES '20*, ACM, <https://doi.org/10.1145/3375627.3375830>, 2020.
- 820 Sohail, T., Zika, J. D., and Ehmen, T. F.: How accurate are salinity measurements around Antarctica? A machine learning based approach, *Machine Learning: Earth*, <https://doi.org/10.1088/3049-4753/ae7113>, 2026.
- Takahashi, T., Sutherland, S. C., Goddard, J. G., and Newberger, T.: Discrete surface measurements of partial pressure of CO₂, dissolved inorganic carbon, nutrients, sea surface temperature and sea surface salinity during the R/V Laurence M. Gould LDEO Drake Passage Surface Water CO₂ Time Series Program from 2002-03-08 to 2017-10-24 (NCEI Accession 0171470), <https://doi.org/10.7289/V5NC5ZGJ>,
 825 2018.
- Talley, L.: Closure of the Global Overturning Circulation Through the Indian, Pacific, and Southern Oceans: Schematics and Transports, *Oceanography*, 26, 80–97, <https://doi.org/10.5670/oceanog.2013.07>, 2013.
- TensorFlow Developers, .: TensorFlow, <https://doi.org/10.5281/ZENODO.4724125>, 2024.
- Terhaar, J., Goris, N., Müller, J. D., DeVries, T., Gruber, N., Hauck, J., Perez, F. F., and Séférian, R.: Assessment of Global Ocean Bio-geochemistry Models for Ocean Carbon Sink Estimates in RECCAP2 and Recommendations for Future Studies, *Journal of Advances in Modeling Earth Systems*, 16, <https://doi.org/10.1029/2023ms003840>, 2024.
- 830 Turner, K. E., Smith, D. M., Katavouta, A., and Williams, R. G.: Reconstructing ocean carbon storage with CMIP6 Earth system models and synthetic Argo observations, *Biogeosciences*, 20, 1671–1690, <https://doi.org/10.5194/bg-20-1671-2023>, 2023.
- Verdy, A. and Mazloff, M. R.: A data assimilating model for estimating Southern Ocean biogeochemistry, *Journal of Geophysical Research: Oceans*, 122, 6968–6988, <https://doi.org/10.1002/2016jc012650>, 2017.
 835
- Wallace, D. W.: Monitoring global ocean carbon inventories, *Ocean Observing System Development*, 1995.
- Weiss, R.: Carbon dioxide in water and seawater: the solubility of a non-ideal gas, *Marine Chemistry*, 2, 203–215, [https://doi.org/10.1016/0304-4203\(74\)90015-2](https://doi.org/10.1016/0304-4203(74)90015-2), 1974.
- Williams, S., Hecht, M., Petersen, M., Strelitz, R., Maltrud, M., Ahrens, J., Hlawitschka, M., and Hamann, B.: Visualization and Analysis
 840 of Eddies in a Global Ocean Simulation, *Computer Graphics Forum*, 30, 991–1000, <https://doi.org/10.1111/j.1467-8659.2011.01948.x>, 2011.
- Zemskova, V. E., He, T.-L., Wan, Z., and Grisouard, N.: A deep-learning estimate of the decadal trends in the Southern Ocean carbon storage, *Nature Communications*, 13, <https://doi.org/10.1038/s41467-022-31560-5>, 2022.